Spring 2020

In Vivo Non-Ribosomal Protein Synthesis in Mammalian Cells

Akhil Choppa

SANTA CLARA UNIVERSITY

Department of Bioengineering

I HEREBY RECOMMEND THAT THE THESIS PREPARED
UNDER MY SUPERVISION BY

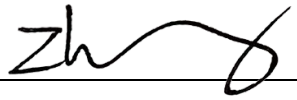
Akhil Choppa

ENTITLED

In Vivo Non-Ribosomal Protein Synthesis in Mammalian Cells

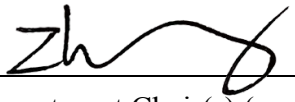
BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE
IN
BIOENGINEERING



June 12, 2020

Thesis Advisor(s) (use separate line for each advisor) Date



June 12, 2020

Department Chair(s) (use separate line for each advisor) Date

In Vivo Non-Ribosomal Protein Synthesis in Mammalian Cells

By

Akhil Choppa

GRADUATE THESIS

Submitted to
the Department of Bioengineering

of

SANTA CLARA UNIVERSITY

in Partial Fulfillment of the Requirements
for the degree of
Master of Science in Bioengineering

Santa Clara, California

Spring 2020

In Vivo Non-Ribosomal Protein Synthesis in Mammalian Cells

Akhil Choppa

Department of Bioengineering
Santa Clara University
2020

ABSTRACT

The mechanism of Sortase A substrate specificity has been widely studied and applied to many approaches to bioconjugation. Current research includes transpeptidation between peptide nucleotide acids, polypeptides, viruses, or antibodies. These applications help improve drug targeting and delivery. Scientists have performed Sortase A-mediated protein ligation *in vitro*. This project proposes an *in vivo* protein ligation method with Sortase A. In this design, Sortase A acts as a catalyst to initiate bioconjugation between the LPETG motif and pentaglycine (Gly5) chain to express GFP. This technique bypasses the ribosome and offers an alternative way to synthesize protein in mammalian cells. Three recombinant plasmids were created with the GFP gene spliced into two segments. The LPETG motif is inserted to the N-terminus of GFP and the Gly5 chain is attached to the C-terminus of GFP. The recombinant plasmids are co-transfected into HeLa cells with another recombinant plasmids that expresses only for Sortase A. The expression of GFP is observed under a fluorescence microscope every 12 hours. Detection of GFP by fluorescence indicates the reconstruction of whole GFP via *in vivo* protein ligation is successful. Overall, this design pioneers synthetic biology by reengineering cellular pathways in a mammalian cell. There is potential to create synthetic cells and apply Sortase A to trigger *in vivo* ligation between a protein drug and a drug target for precision medicine.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Roshani B. Cowmeadow for providing the fourth recombinant that encodes only for Sortase A. Roshani was a previous postdoc student in Dr. Zhang's lab. She created this recombinant plasmid that was used for the experimental designs in this project.

Next, I thank Dr. Zhiwen Zhang for providing the funding and lab space for this research. He is the principal investigator of this project. Dr. Zhang taught me the background of Sortase A and guided me throughout my master's program. He mentored me on biomolecular lab techniques such as mammalian & bacterial cell culture, protein purification, DNA extraction, and transformation/transfection.

Finally, I would like to thank the students and peers who helped me throughout this project. Ben and Sarah provided extra hands in the lab to carry out the experient. Their input on the data was helpful and I appreciate their feedback on the results.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER 1 - BACKGROUND

Sortase A is a protein that is expressed by gram-positive bacterial cells. This protein has a molecular weight of 27 kDa. It was first discovered in 1999 and ever since, Sortase A has been widely studied [1]. Sortase A is responsible for anchoring surface proteins on the membrane and activating the infection by gram-positive bacterial cells [1]. For infectious gram-positive bacteria, the outer membrane of the bacterial cell is covered with chains of glycans that form elastic networks and protect the bacterial cell from lysis. The glycan chains are about 20 to 30 disaccharide units in length and are linked together by covalent bonds [2]. Gram-positive bacteria are harmful and pathogenic. Gram-positive bacteria can cause skin infections, pneumonia, and other inflammatory diseases by invading mammalian cells [3].

Today, Sortase A is applied to other applications such as directed evolution, perform site-specific protein labeling, and enhance other transpeptidase ligases [4, 5]. In this project, the mechanism of Sortase A is used as protein ligase for *in vivo* protein synthesis in mammalian cells. The *in vivo* protein ligation will reengineer the cellular network for protein synthesis by bypassing the ribosomes. This will revolutionize synthetic biology and help understand other diverse cellular pathways and processes [6, 7]. Also, *in vivo* protein ligation provides the best selectivity of drug targeting that can be beneficial in precision medicine.

1.1 SORTASE A SUBSTRATE SPECIFICITY

The mechanism of Sortase consists of cross-linking between multiple bacterial membrane peptides [8]. **Figure 1** portrays an image of the reaction that anchors the proteins on the bacterial surface and activates gram-positive. Initially, a surface protein precursor of Sortase A is created with a sorting signal and LPXTG motif [9]. The sorting signal is located on the N-terminus and the LPXTG motif is located on the C-terminus of this surface protein precursor. The sorting signal guides the surface protein precursor to the secretory pathway. The secretory pathway recognizes the sorting signal and removes it [10]. The remaining segment of the surface protein precursor is incorporated on the bacterial membrane. Sortase A proteins on the surface of the bacterial cell recognize the LPXTG motif of the surface protein precursor and cleaves between the threonine and glycine [10]. This cleavage creates a carboxyl group on the threonine.

Another peptide group known as Lipid II is located on the bacterial cell surface. Lipid II has a pentaglycine (Gly5) chain that interacts with the carboxyl group on the threonine and forms an amide bond [11]. **Figure 1** displays a visualization of the splice site between the threonine and glycine on the LPXTG motif. Also, it shows the cross-link between the threonine and the Gly5 chain. The remaining segment of the surface protein precursor is incorporated into the bacterial cell membrane. This cell wall incorporation is completed via a transglycosylation reaction [12]. The surface protein precursor is incorporated on the cell membrane of gram-positive bacteria. Wild-type Sortase A is a dimer protein, however scientists have performed site-directed mutagenesis to create a mutant monomeric form of Sortase A [13].

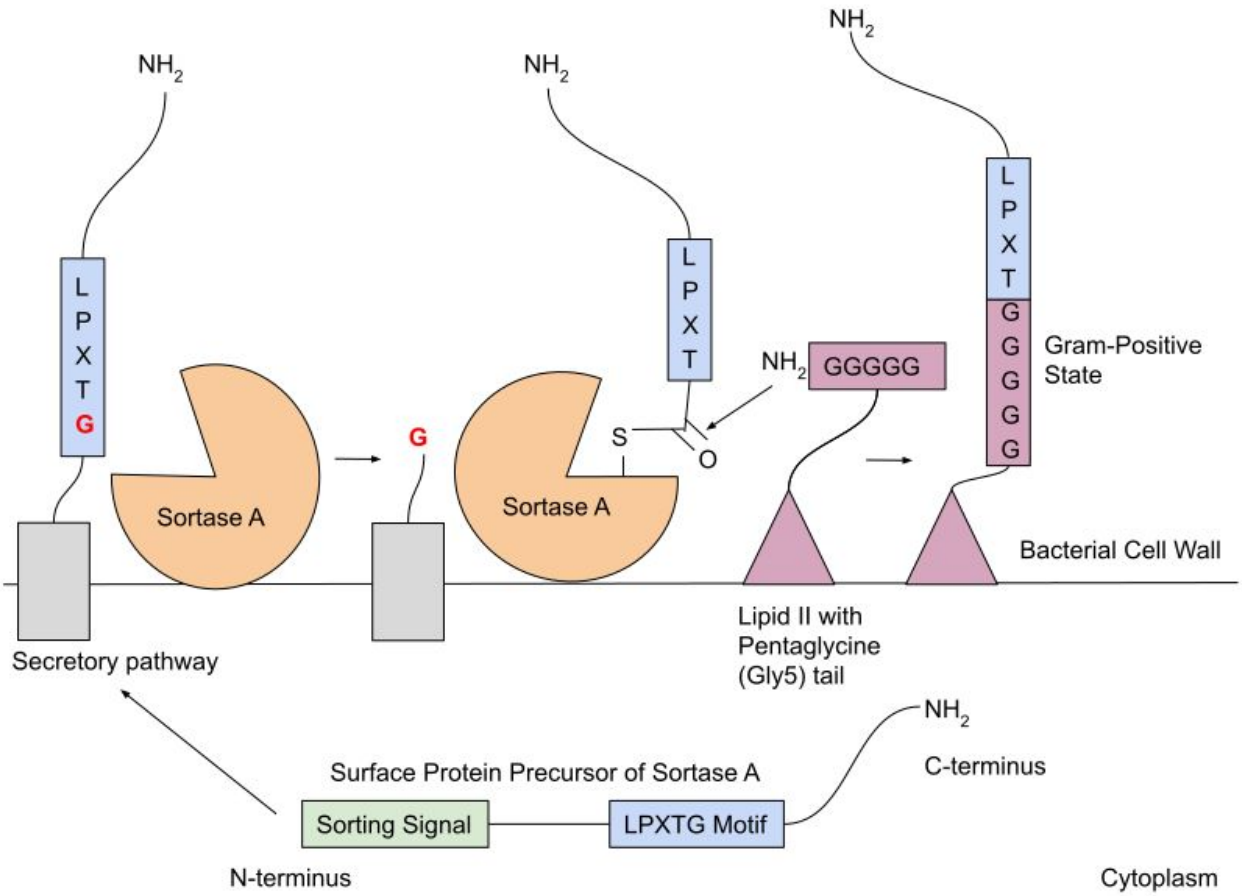


Figure 1. Illustration of the mechanism of Sortase A in gram-positive bacterial cells. A surface protein precursor of Sortase A is created with a sorting signal on N-terminus and LPXTG motif on C-terminus. The sorting signal guides the surface protein precursor to the secretory pathway. The secretory pathway recognizes the sorting signal and removes it. Sortase A identifies the LPXTG motif and cleaves in between the threonine and glycine. The cleavage creates a carboxyl group on the threonine that interacts with the pentaglycine (Gly5) chain on Lipid II. This cross-link forms an amide bond and incorporates the remaining surface protein precursor into the bacterial cell wall. At this point, the gram-positive bacteria is infectious.

1.2 CLICK CHEMISTRY

Click chemistry is a popular bioconjugation tool to link substrates together. This method has multiple applications such as drug discovery in medicinal chemistry and drug development in pharmaceutical sciences [14]. Click chemistry is used in research to link biocompatible polymers via small molecular reactions. There are studies on bioconjugation of various nanoparticles to create complex structures for drug delivery and selective labeling [15, 16]. For example, carbon nanoparticles can bind to each other shapes like micelles, nanotubes, nanospheres, and nanocapsules that serve as alternatives to viral delivery systems. Also, fluorescent, gold, and magnetic nanoparticles can act as markers for labeling antibodies, DNA, cells and other small molecules. This is especially helpful in cancer treatment, since tumor cells can be identified more easily [17]. There are some commercially available drugs that were

created with click chemistry. These drugs are possible cancer therapies that conjugate nanoparticles to antibodies, known as antibody drug conjugates (ADC) [18]. Click chemistry is easy to perform, since it can occur *in vivo* or *in vitro* with the same effects. The reaction will not interfere or interact with other native biochemical processes. The reaction can take place in water and other mild nontoxic environments [16]. Some of the most common click chemistry reactions are copper-catalysed azide-alkyne cycloaddition (CuAAC) triazole annulation and sulfur fluoride exchange (SuFEx) catalysis [19].

Click chemistry has many purposes, however there are some limitations and side effects. Click chemistry reactions use copper and platinum as catalysts. Copper and platinum are toxic compounds that cause cell death. Due to this, click chemistry can not be used *in vivo*. There is a need to transition from click chemistry to a more natural bioconjugation method.

1.3 BIOCONJUGATION WITH SORTASE A

The function of Sortase A has been studied and applied as an alternative approach to bioconjugation. Sortase A is used as a catalyst to conjugate polypeptides together. This method is better than click chemistry, since it does not require any toxic compounds. Currently, there is research with Sortase A-mediated ligation [20]. The Sortase A mechanism is used to bind peptide nucleic acids with other peptides, antibodies, or viruses [21, 22]. The goal of this is to create cell-penetrating peptide virus chimeras and antibody drug conjugates that can deliver cytotoxic drugs to cancer cells more efficiently [21, 22, 23]. Sortase A conjugation is applied to the N-terminus and C-terminus of both the heavy and light chains of the anti-HER2 antibody to create multiple variants [23]. The substrate specificity between the LPXTG motif and pentaglycine chain is implemented in joining nanoparticles structures together and combining functional forms of DNA-protein nanostructures [24]. In addition, the Sortase A enzyme and substrate specificity is engineered to improve the efficiency of transpeptidation. The variants of Sortase A enchance N-terminus and C-terminus labeling for cell surface modification [25, 26].

Other research includes directed evolution of the LPXTG motif. This sorting motif was altered into other substrates like LAXTG and LPXSG that still maintain a high activity and specificity with Sortase A [27]. Even with all these advancements, there are some limitations with Sortase A-mediated protein ligation. Transpeptidations via Sortase A have been performed in test tubes and on the surface of living cells [28]. There are no applications of Sortase A protein ligations *in vivo*. Bioconjugation with Sortase A requires expression and posttranslational modifications to catalyze the interaction between the LPXTG motif and pentaglycine chain [29]. Thus, the majority of research with Sortase A-mediated ligation is performed *in vitro*. In order for bioconjugation with Sortase A to take place *in vivo*, the LPXTG motif, pentaglycine chain, and Sortase A enzyme need to be expressed successfully in mammalian cells [30]. **Figure 2** illustrates a proof of concept to perform Sortase A-mediated protein ligation *in vivo*. The Sortase A enzyme is expressed separately from the LPXTG motif and pentaglycine chain. In other words, two different recombinant plasmids need to be expressed *in vivo*. One recombinant

plasmid encodes for the Sortase A gene while the second recombinant plasmid encodes for the LPXTG motif and pentaglycine chain. This way, Sortase A will carry out transpeptidation in mammalian cells.

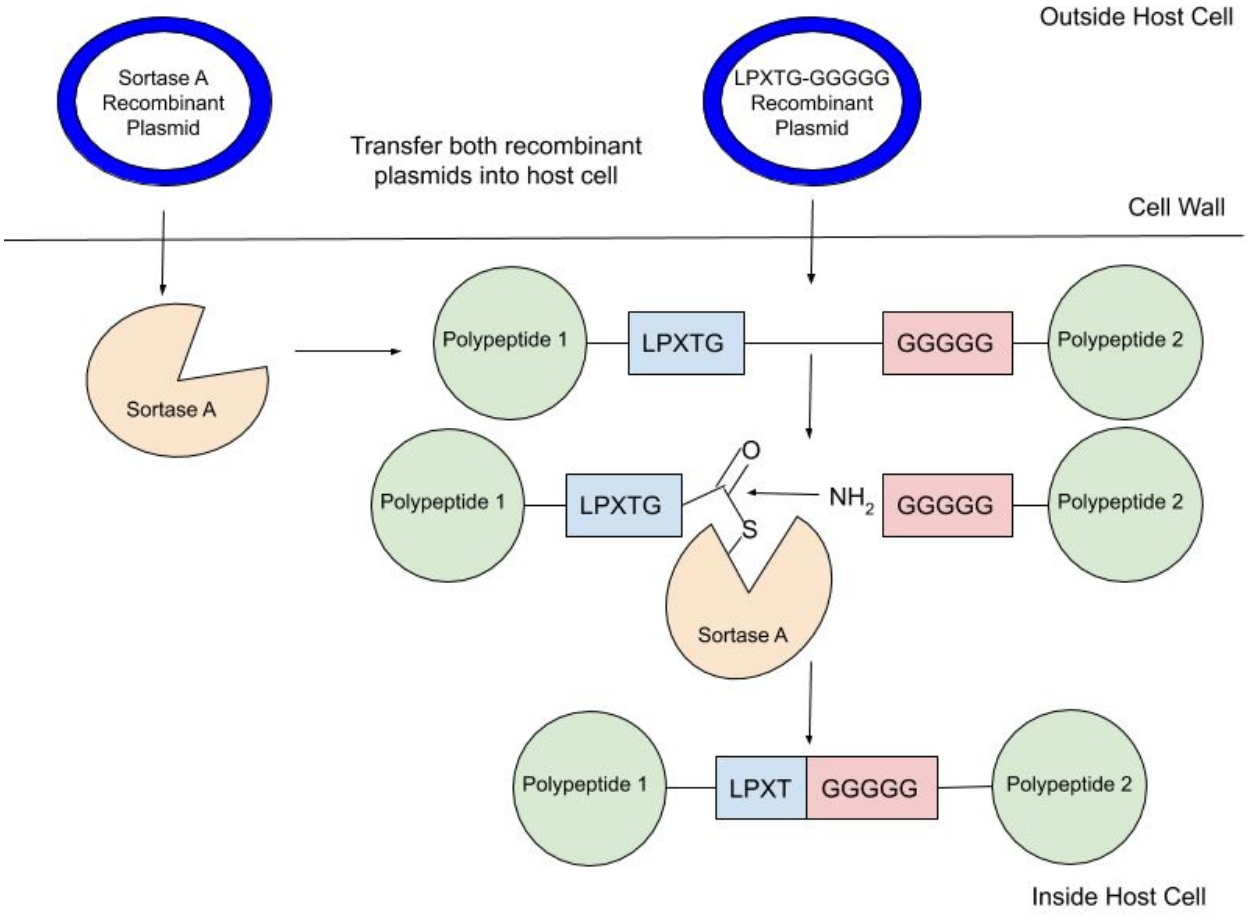


Figure 2. Diagram of Sortase A-mediated protein ligation *in vivo*. Current research of bioconjugation via Sortase A is performed *in vitro*. This diagram displays a proposal of using Sortase A to catalyze transpeptidation between the LPXTG motif and pentaglycine (Gly5) chain *in vivo*. Sortase A and LPXTG-GGGGG are encoded in two separate recombinant plasmids. Both recombinant plasmids are transferred into a host cell and expressed. When Sortase A is expressed, it will cleave between the threonine and glycine on the LPXTG motif. Then, the remaining motif of the surface protein precursor and pentaglycine chain will conjugate together.

1.4 SYNTHETIC BIOLOGY

Synthetic biology is a popular upcoming field in science, because it focuses on the minimality, modularity, and controllability of complex cellular functions [31]. Scientists have investigated the diverse network and metabolic pathways of a mammalian cell. The concept of artificial cells was first introduced in 1957. Its purpose is to mimic the same genetic circuitry of a natural cell. Ever since then, synthetic biology was applied to replicate the components of the cell. Artificial cell membranes were constructed with lipids, fatty acids, polymers [32]. Tools such as alternative splicing, RNAi, and epigenetics help understand cellular mechanisms. These methods can make complex biological functions more reliable, efficient, and predictable [31]. In

order to create artificial cells, scientists have modified the genome to only the essential genes to carry out cell signaling, cell communication, cell cycle, and differentiation. Another approach was to start from non-living components and assemble them together into an artificial cell [33]. Also, synthetic biology uses cellular components like transcription factors to reroute signaling pathways. These applications still have some challenges and limitations.

Artificial cells have the potential to create artificial organs and tissue for transplantations. They can be used as the next drug delivery mechanism to treat diseases [34]. However, these applications are performed *in vitro*. There are high chances that artificial cells will get rejected by the immune system when introduced *in vivo* [35]. This project pioneers a new method in synthetic biology by reengineering natural cells with novel proteins and functions. Sortase A is introduced into mammalian cells to catalyze *in vivo* protein synthesis. Therefore, this mechanism bypasses the use of ribosomes. In other words, this experiment reengineers the cellular network to create synthetic mammalian cells with new functions [36]. The benefits of this method are biocompatibility with the immune system and exclusion of complex biological functions like transcription and translation.

CHAPTER 2 - DESIGN

The goal is to demonstrate *in vivo* protein ligation in mammalian cells with the presence of Sortase A and completely bypass ribosomal interaction. Sortase A is a protein expressed in gram-positive bacteria. This activation is initiated by the interaction between a cleaved LPETG motif and pentaglycine (Gly5) chain from lipid II to form cell wall incorporation. This interaction is utilized in a recombinant plasmid system and transfected into HeLa cells to observe *in vivo* protein ligation.

2.1 RECOMBINANT PLASMIDS

Three recombinant plasmids were designed to encode for the protein of interest using GFP tag to verify transfection. Each recombinant plasmid was designed to express a different level of GFP. These three recombinant plasmids would be co-transfected into HeLa cells with a fourth recombinant that encodes for Sortase A. The purpose of this is to express Sortase A separately from the LPETG motif and Gly5 chain. This way, once Sortase A is introduced, *in vivo* ligation between the LPETG motif and Gly5 chain will occur.

Plasmid 1 is the original design with the LPETG motif and Gly5 chain. The GFP gene is cleaved into two segments. The LPETG motif is inserted directly after the N-terminus of GFP and the Gly5 chain is inserted before the C-terminus of GFP. Plasmid 2 includes the addition of TOP1 and SH3 domains. SH3 domain is involved in many important regulatory pathways responsible for cytoskeletal growth and cell proliferation [37]. There are high-affinity peptides that bind to the SH3 domain. One of these domains is TOP1 that has a high binding affinity with SH3 [38]. The TOP1 domain is inserted in between the N-terminus of GFP and the LPETG

motif. While the SH3 domain is inserted in between the Gly5 chain and the C-terminus of GFP. Plasmid 3 includes the addition of BOT1 and SH3 domains as negative control. The BOT1 domain has no binding affinity with SH3 [39, 40]. The BOT1 domain is inserted in between the N-terminus of GFP and the LPETG motif. While the SH3 domain is inserted in between the Gly5 chain and the C-terminus of GFP. TOP1 and BOT1 domains were tested with SH3 and they show differences in binding affinities when forming covalent linkage to a target protein [40]. Therefore, TOP1 and BOT1 were implemented into plasmid 2 and plasmid 3 respectively to verify the specificity of *in vivo* protein ligation via Sortase A. **Figure 3** shows a schematic of the three recombinant plasmids designed for this study.

Figure 3. Schematic of the three recombinant plasmids. (**A**) Represents plasmid 1 that includes LPETG motif on the N-terminus of GFP and Gly5 chain on C-terminus of GFP. (**B**) Represents plasmid 2 that includes TOP1 domain and LPETG motif on N-terminus of GFP and Gly5 chain and SH3 domain on C-terminus of GFP. (**C)** Represents plasmid 3 that includes BOT1 domain and LPETG motif on N-terminus of GFP and Gly5 chain and SH3 domain on C-terminus of GFP.

The goal is to perform *in vivo* Sortase A-mediated bioconjugation by bringing the N-terminus and C-terminus of GFP in proximity and allow protein ligation. Plasmid 4 that encodes only for Sortase A is introduced into HeLa cells with either plasmid 1, plasmid 2, or plasmid 3. Plasmid 1 expressess the N-GFP-LPETG and Gly5-C-GFP. Plasmid 2 expresses the N-GFP-LPETG-TOP1 and SH3-GLY5-C-GFP. TOP1 and SH3 domains have a high binding affinity with each other [38]. Thus, this will allow a very high expression of GFP. Plasmid 3 expresses N-GFP-LPETG-BOT1 and SH3-GLY5-C-GFP. BOT1 and SH3 domains have no binding affinity with each other [39]. Therefore, this will lead to a very low signal due to the autofluorescence of mammalian cells.

2.2 PROTEIN EXPRESSION

HeLa cells were chosen as the mammalian cell system to observe protein expression. They are cultured to a confluency of 90% and transfected with plasmid 4 and one of plasmids 1-3. **Figure 4** displays a diagram of the transfection layout with the different recombinant plasmids After transfecting the recombinant plasmids into the HeLa cells, expression of GFP will be observed under a fluorescence microscope every 12 hours. If fluorescence is detected, then

this indicates that Sortase A triggered bioconjugation between the LPXTG motif and Gly5 chain. When this interaction happens, the N-terminus and C-terminus of GFP will ligate back together reconstructing the whole GFP and giving off fluorescence.

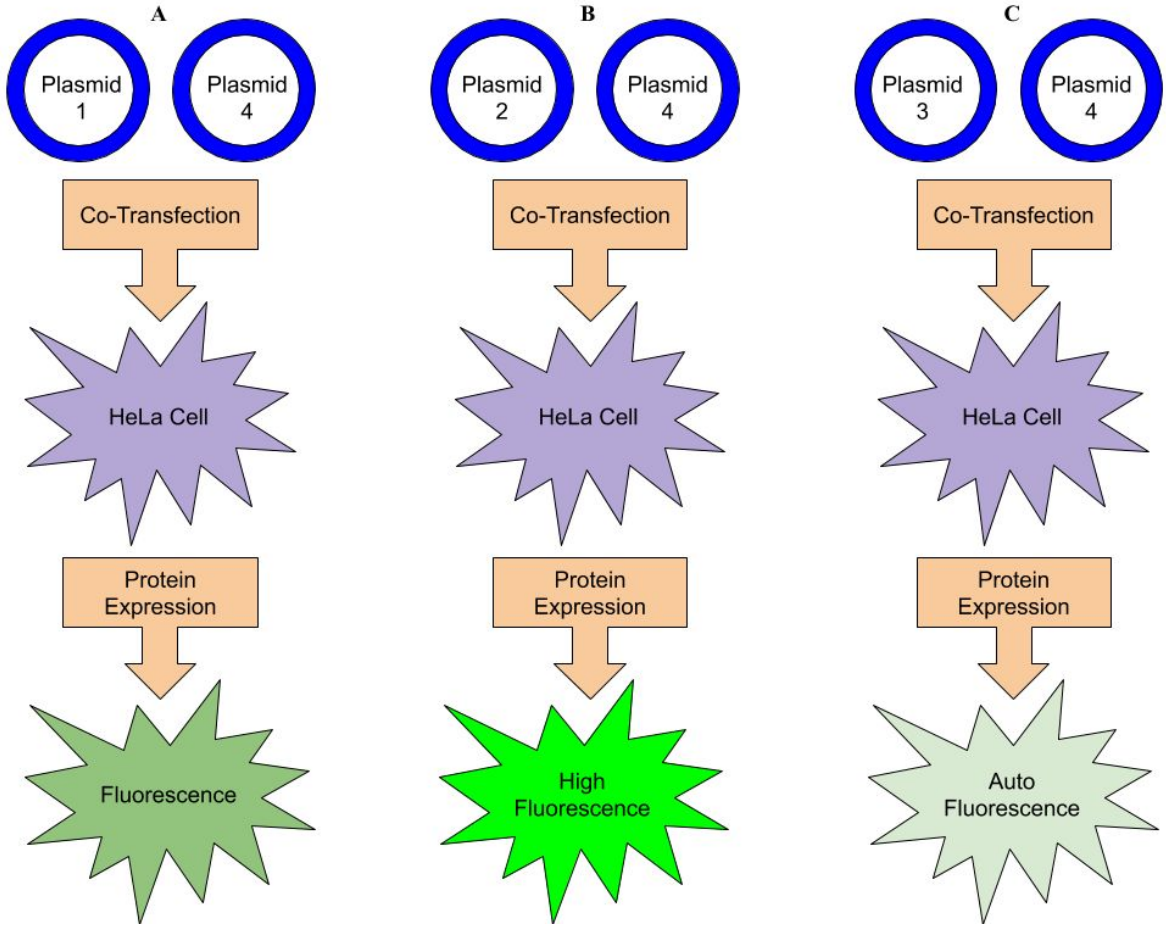


Figure 4. Layout of transfection with designed recombinant plasmids. (**A**) Plasmid 1 and plasmid 4 are both transfected into HeLa cells. Protein expression indicates *in vivo* protein ligation. This is visually proven when the HeLa cells show a green fluorescence. (**B**) Plasmid 2 and plasmid 4 are both transfected into another set of HeLa cells. This combination will show intense fluorescence due to the high binding affinity between the TOP1 and SH3 domains in plasmid 2. (**C**) Plasmid 3 and plasmid 4 are both transfected into a third batch of HeLa cells. Because plasmid 3 has no binding affinity between the BOT1 and SH3 domains, it will result in autofluorescence.

CHAPTER 3 - METHODS AND MATERIALS

The three recombinant plasmids were designed with a software called ApE (version 2.0.61). The fourth recombinant plasmid was already previously created. Once the recombinant plasmids were created and optimized for mammalian expression, they were transformed into One Shot™ TOP10 competent cells from Invitrogen for amplification. Afterwards, the recombinant plasmids were extracted and isolated with Qiagen miniprep and midiprep kits. At the same time, HeLa cells were cultured to a confluency of 90%. The HeLa cells were transfected with either one of the following groups of recombinant plasmids: plasmid 1 & plasmid 4, plasmid 2 &

plasmid 4, plasmid 3 & plasmid 4. After transfection, the HeLa cells were monitored under a fluorescence microscope every 12 hours for protein expression.

3.1 CREATING RECOMBINANT PLASMIDS

The pEF-GFP vector plasmid (catalog #11154) from Addgene is a commercially available vector backbone that was used for creating the recombinant plasmids. The GFP gene is included in the pEF-GFP vector plasmid. This vector plasmid was selected, because it is small with 3086 nucleotides and expresses well in a mammalian cell system. The pEF-GFP vector plasmid has ampicillin as the bacterial resistance, origin of replication (ori) region, and EF1alpha promoter.

Plasmids 1-3 were designed on ApE to show *in vivo* protein ligation of GFP. The DNA sequence of GFP was retrieved from SnapGene (version 5.1). The entire DNA sequence of GFP is shown in Appendix A. The GFP sequence was cleaved into two segments. The DNA sequence was cleaved in between the 471st and 472nd base pairs. In other words, the GFP will ligate back together at the 157th and 158th amino acids. This cleavage site was selected, because it was proven that the native form of GFP can be spliced at this location, modified, and ligated back together to restore protein function [41]. Due to this, GFP can be cleaved into two segments and link back together without losing protein function.

The LPETG motif was inserted on the N-terminus of GFP and Gly5 chain was inserted on the C-terminus of GFP. Other components such as a kozak sequence, FLAG sequence, start codon, stop codon, and polyadenylation signal were added too. **Figure 5** is a vector map of plasmid 1. When Sortase A is expressed in HeLa cells with plasmid 1, Sortase A will splice between the threonine and glycine on the LPETG motif. The carboxyl group on the threonine will interact with the Gly5 chain and bind together. This linkage brings the N-terminus and C-terminus of GFP together to express the protein. The kozak sequence is used to indicate the beginning of the N-terminus and C-terminus of GFP. The FLAG sequences are placed directly before the LPETG motif and after Gly5 chain. They are used to label and identify when the recombinant plasmid is expressed. A start codon (ATG) is added directly after the kozak sequence on the N-terminus and C-terminus of GFP to transcribe both segments of GFP. A stop codon (TAA) is placed after the LPETG motif and C-terminus of GFP to end transcription. Finally, a polyadenylation signal is added after both stop codons to stabilize the mRNA sequences. The complete sequence of plasmid 1 is Appendix B.

Other modifications were added to plasmid 2 and plasmid 3. Plasmid 2 and plasmid 3 were designed to have different levels of GFP expression. Plasmid 2 and plasmid 3 have the same components as plasmid 1, however plasmid 2 includes TOP1 and SH3 domains, while plasmid 3 includes BOT1 and SH3 domains. For plasmid 2, the TOP1 domain is inserted directly before the LPETG motif on the N-terminus of GFP and the SH3 domain is inserted in between the FLAG sequence and C-terminus sequence of GFP. A vector map of plasmid 2 is displayed in **Figure 6**. Appendix C shows the full-length sequence of plasmid 2. The purpose of adding the

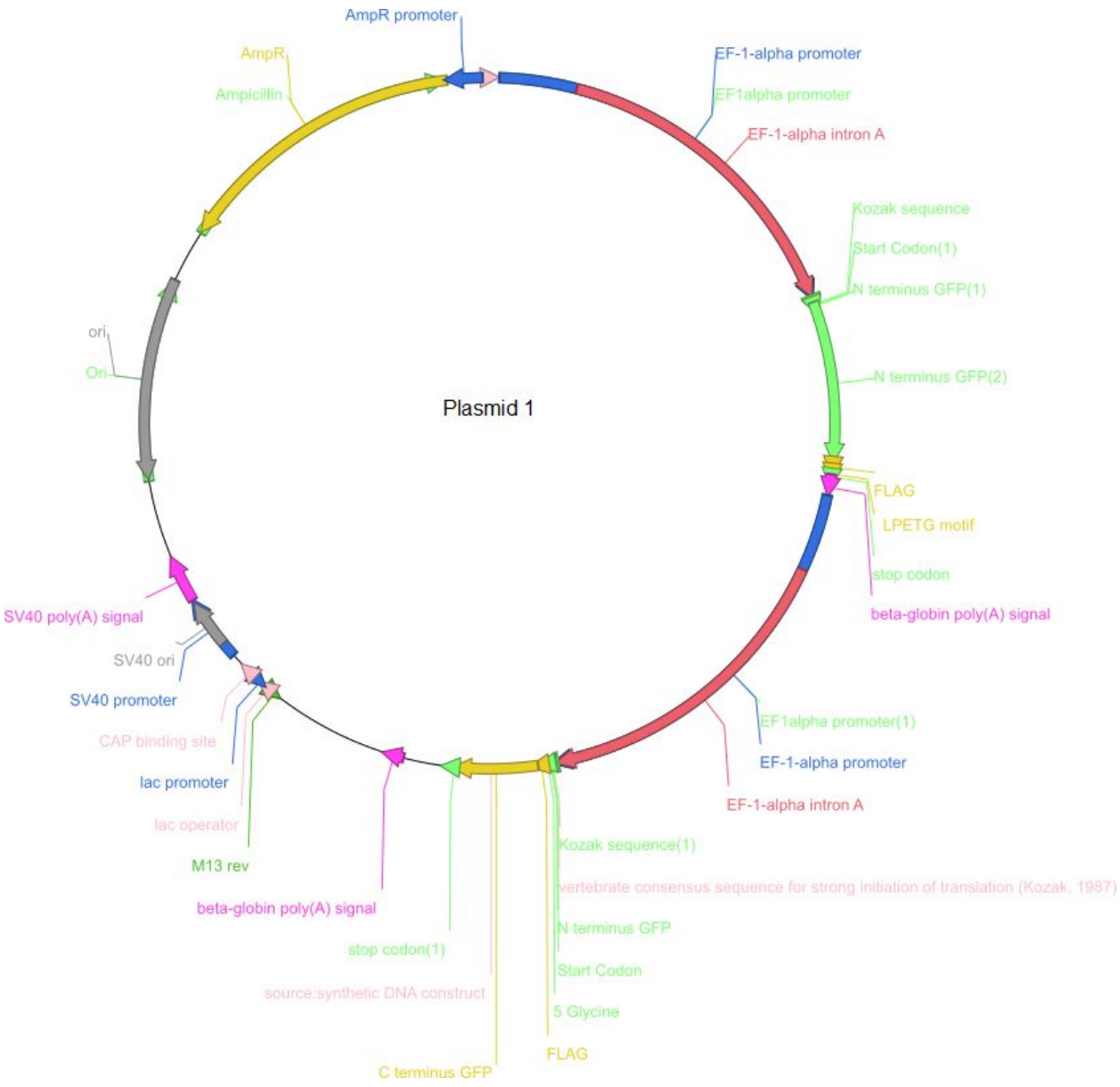


Figure 5. Vector map of plasmid 1. This is the original recombinant plasmid design. The GFP gene is spliced into two segments. N-terminus of GFP is 471 nucleotides and C-terminus of GFP is 245 nucleotides. The LPETG motif is inserted on the N-terminus of GFP while the Gly5 chain is added on the C-terminus of GFP. This recombinant plasmid includes EF1alpha promoter, origin of replication (ori), ampicillin as bacterial resistance, kozak sequence, FLAG sequence, start codon (ATG), stop codon (TAA), and polyadenylation signal.

TOP1 and SH3 domain is to increase the expression of GFP. TOP1 and SH3 domains have a high binding affinity to each other [38]. Therefore in the presence of Sortase A, the LPETG motif and Gly5 chain will interact and bring the N-terminus and C-terminus of GFP segments together. During this process, the TOP1 and SH3 domains will link together too [40]. Plasmid 2

has two methods to trigger protein ligation. Thus, N-terminus and C-terminus of GFP are more likely to bind together and result in a higher protein expression.

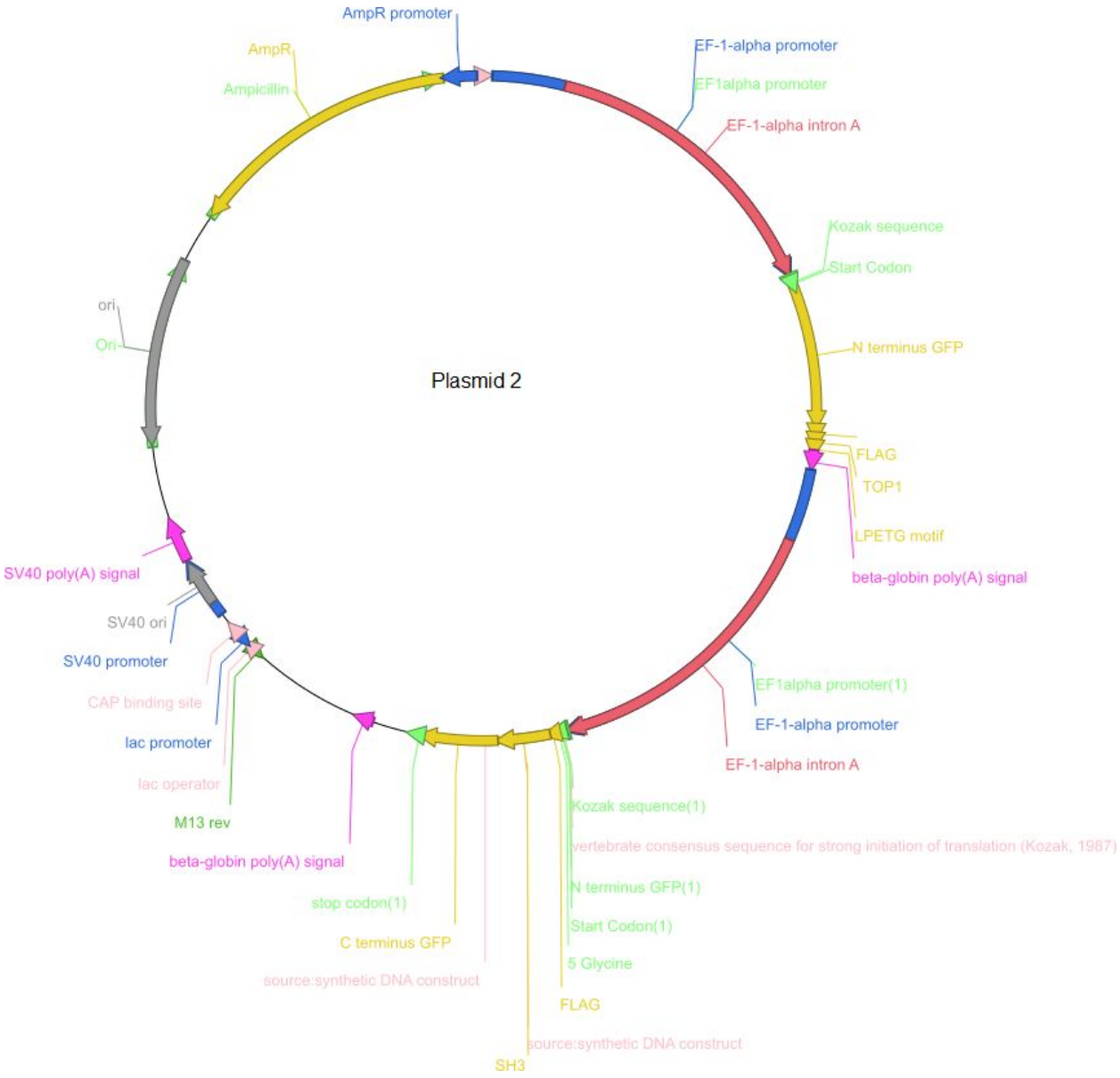


Figure 6. Vector map of plasmid 2. This recombinant plasmid has the same features as plasmid 1 with some additional components. It includes GFP gene spliced into two segments, LPETG motif, Gly5 chain, EF1alpha promoter, origin of replication (ori), ampicillin as bacterial resistance, kozak sequence, FLAG sequence, start codon (ATG), stop codon (TAA). and polyadenylation signal from plasmid 1. The additional elements include TOP1 and SH3 domains. These domains have a high binding affinity with each other. Therefore, this recombinant plasmid will have a higher protein expression of GFP.

Plasmid 3 has BOT1 and SH3 domains that have no binding affinity to each other [39]. When Sortase A is introduced to plasmid 3, the LPETG motif and Gly5 chain will link together. However, the BOT1 and SH3 domains will prevent protein ligation between the N-terminus and C-terminus of GFP [40]. Overall, plasmid 3 will cause a low protein expression. **Figure 7** presents a schematic of plasmid 3 and the complete sequence is in Appendix D.

Figure 7. Vector map of plasmid 3. This recombinant plasmid has the same features as plasmid 1 with some additional components. It includes GFP gene spliced into two segments, LPETG motif, Gly5 chain, EF1alpha promoter, origin of replication (ori), ampicillin as bacterial resistance, kozak sequence, FLAG sequence, start codon (ATG), stop codon (TAA). and polyadenylation signal from plasmid 1. The additional elements include BOT1 and SH3 domains. These domains do not have a binding affinity with each other. Therefore, this recombinant plasmid will result in autofluorescence of HeLa cells.

Plasmid 4 that contains the gene that encodes for Sortase A. This recombinant plasmid was previously created and stored in -80°C. The Sortase A gene is 621 base pairs long and the protein structure is made up of 206 amino acids. The recombinant plasmid is 710 nucleotides in total [9, 13]. More information about plasmid 4 is available in Appendix E. Plasmid 4 is paired with each recombinant plasmid to initiate protein ligation and signal protein expression.

3.2 PREPARATION OF RECOMBINANT PLASMIDS

Once the recombinant plasmids were created, they were replicated and amplified in bacterial cells. One Shot™ TOP10 competent cells (catalog #404010) from Invitrogen were used for transformation. Each recombinant plasmid was transformed into separate reactions. Each reaction has a transformation efficiency of 1 x 10^9 cfu/µg plasmid DNA. One Shot™ TOP10 competent cells have a high cloning efficiency, efficient plasmid propagation rate, and stable DNA replication. The recombinant plasmids were added in different reactions of One Shot™ TOP10 competent cells and incubated on ice for 30 minutes. Then, each reaction was heated-shocked for 30 seconds at 42°C without shaking and immediately placed on ice for 2 minutes. Prewarmed S.O.C medium (catalog #155544034) from Thermo Fisher Scientific was added to each reaction aseptically and placed in a 37°C shaking incubator at 225 rpm for 1 hour. LB agar plates were prepared with 50 µg/mL Ampicillin under sterile conditions. Once the LB agar plates solidified and cooled down to room temperature, each transformation reaction was spread onto a separate LB agar plate. The transformed LB agar plates were inverted and stored in an incubator at 37°C overnight. After transformation, the recombinant plasmids were isolated and extracted with miniprep (catalog #27104) and midiprep (catalog #12143) from Qiagen.

Transformed colonies formed on the LB agar plates overnight. For miniprep, the Qiagen protocol was followed. A single colony was picked from each plate and inoculated in 5 mL autoclaved LB broth media with 50 µg/mL Ampicillin. Each inoculation was incubated overnight at 225 rpm and 37°C. Each inoculation was transferred to a seperate large scale LB broth media of 75 mL. Each bacterial culture was placed in a 37°C shaking incubator at 225 rpm for 4 hours. The OD of plasmid 1, plasmid 2, plasmid 3, and plasmid 4 were 0.548, 0.489, 0.668, and 0.735 respectively. Each bacterial culture was harvested by centrifuging at 6000 rpm for 15 minutes at 4°C. Each bacterial pellet was resuspended in 1.5 mL of Buffer P1. Then 1.5 mL of buffer P2 was added to each resuspension and vigorously inverted. Each resuspension incubated at room temperature for 5 minutes. Then 2.1 mL of Buffer P3 was added to each resuspension. Each resuspension was inverted again and incubated on ice for 5 minutes. Each resuspension was centrifuged for 10 minutes at 16,000 rpm. 0.5 mL of Buffer PB was pipetted into separate QIAprep 2.0 spin columns and centrifuged for 1 minute. Then 800 µL of each supernatant was pipetted into the QIAprep 2.0 spin columns and spun down for 1 minute. 0.75 mL of Buffer PE was pipetted to each QIAprep 2.0 spin column and centrifuged for 1 minute to wash the columns. The flowthrough was disposed and each QIAprep 2.0 spin column was spun down again for another minute to remove any remaining residual wash buffer. To elute the recombinant

plasmids, 50 µL of Buffer EB was pipetted into each QIAprep 2.0 spin column and centrifuged for 1 minute. The flowthrough was collected and measured for concentration on a Nanodrop. The concentration of plasmid 1, plasmid 2, plasmid 3, and plasmid 4 were 62.9 ng/µL, 42.8 ng/µL, 65.1 ng/µL, and 139.7 ng/µL respectively.

For midiprep, the Qiagen protocol was followed as well. Two single colonies were picked from each plate and inoculated into two separate 5 mL autoclaved LB broth media with 50 µg/mL Ampicillin. The small scale cultures were placed in a shaking incubator overnight at 225 rpm and 37°C. Afterwards, two of the same inoculations were transferred to 150 mL of large scale LB broth media. Each large scale bacterial culture incubated for 5 hours at 225 rpm and 37°C. The OD of plasmid 1, plasmid 2, plasmid 3, and plasmid 4 were 0.803, 0.772, 0.768, and 0.519 respectively. Each large scale bacterial culture was harvested by centrifugation at 6000 rpm for 15 minutes at 4°C. Then, the bacterial pellet was resuspended in 4 mL of Buffer P1. 4 mL of Buffer P2 was added to each resuspension and mixed thoroughly. The resuspensions incubated for 5 minutes at room temperature. Then, 4 mL of Buffer P3 was added to each resuspension and mixed thoroughly again. The resuspensions incubated for 15 minutes on ice. Each resuspension was centrifuged at 20,000 for 30 minutes at 4°C. Qiagen-tips were equilibrated with 4 mL of Buffer QBT and emptied by gravity flow. After centrifugation, the supernatant from each resuspension was pipetted into a seperate Qiagen-tip. The supernatant flowed through the column by gravity. To wash each Qiagen-tip, 10 mL of Buffer QC was pipetted and flowed through the column by gravity. The flowthrough was discarded and 5 mL of Buffer QF was pipetted to each Qiagen-tip to elute the recombinant plasmid. Again, gravity force emptied the column and the flowthrough was collected. The concentration of each recombinant plasmid was measured on a nanodrop. The concentration of plasmid 1, plasmid 2, plasmid 3, and plasmid 4 were 43.0 ng/µL, 44.5 ng/µL, 35.5 ng/µL, and 34.9 ng/µL respectively.

3.3 CULTURING HELA CELLS

HeLa cells were selected for the mammalian expression system, because they are more robust and capable of surviving multiple passages. HeLa cells tend to maintain their shape and function even after ten passages [42]. Also, HeLa cells have a longer growth rate in between passages before mutating. Therefore, they do not need to be passaged very frequently [43, 44]. For this study, the HeLa cells were passaged every 48 hours. The HeLa cells were prepared in a specific way for transfection. **Figure 8** illustrates the layout of how the HeLa cells were passaged for transfection. First, the cell media that consisted dMEM (catalog #10566016) and 10% FBS (catalog #16140063), PBS (catalog #10010023), and Trypsin (catalog #25200056) from Thermo Fisher Scientific were preheated. The HeLa cells were washed in 1.5 mL of PBS twice. To detach the HeLa cells, 1 mL of Trypsin was added and incubated at 37°C for 2 minutes. 3 mL of cell media was added and detached HeLa cells were centrifuged for 1 minute at 1500 rpm. The pellet was resuspended in 3 mL of cell media. A 6-well lab plate was prepared with 2 mL of cell media is each well. 1 mL of the resuspended cell pellet was pipetted into each well of the 6-well

lab plate. HeLa cells cultured on a single 10 mL lab plate could be passaged into three individual wells of a 6-well lab plate. Initially, HeLa cells were cultured on three 10 mL lab plates for 48 hours. After 48 hours, the cell confluency was about 80%. For transfection, the HeLa cells from the three 10 mL lab plates were passaged into seven individual wells of a 6-well lab plate. The remaining 2 mL of cell resuspension was passaged into another three 10 mL lab plates. Each 10 mL lab plate received 660 µL of the remaining cell resuspension. The new set of 10 mL lab plates were cultured in a 37°C incubator for the next transfection.

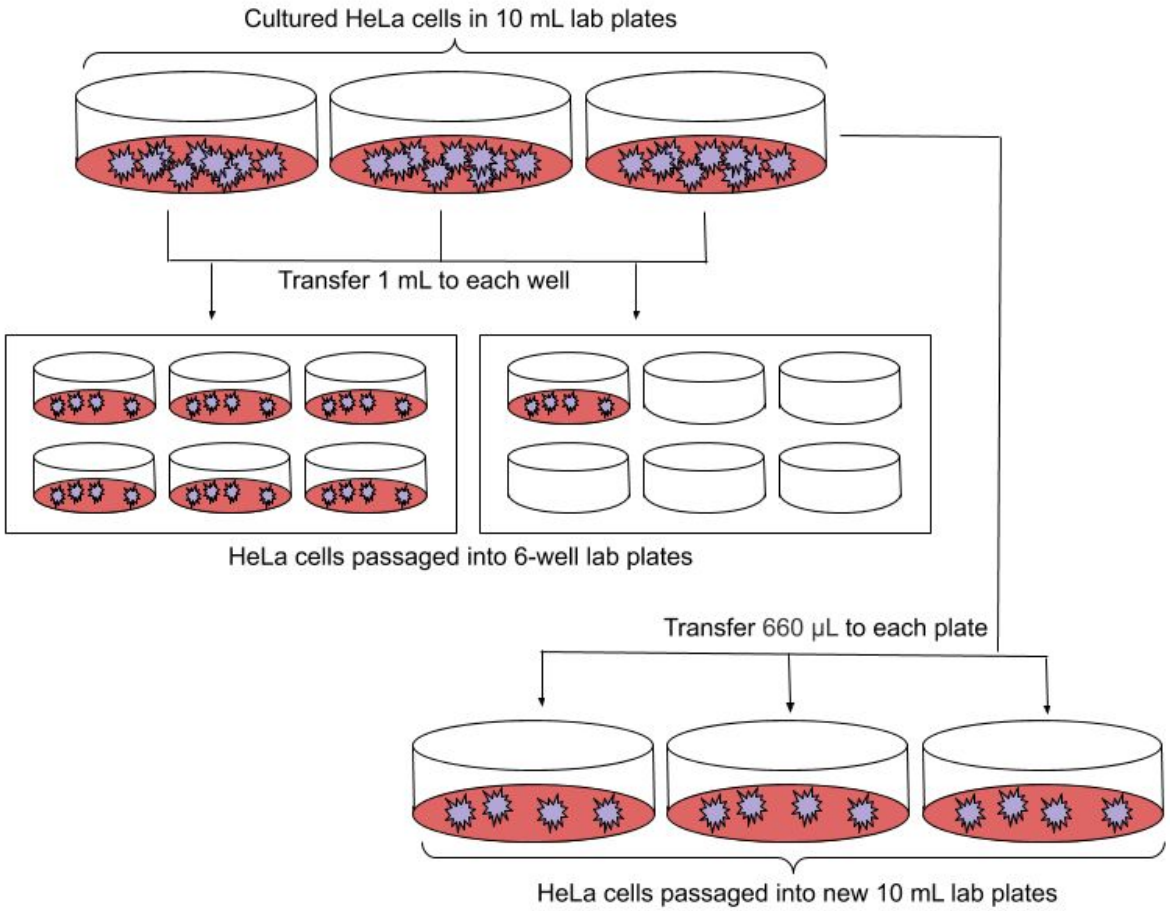


Figure 8. HeLa cell passage layout for transfection. Initially, HeLa cells are cultured on three 10 mL lab plates and grown for 48 hours in a 37°C incubator. Once the confluency is about 80%, the HeLa cells are passaged into a 6-well lab plate for transfection. Each 10 mL lab plate is enough to passage HeLa cells into three wells of a 6-well lab plate. 1 mL of HeLa cells are passaged into each well. In total, seven wells of 6-well lab late are prepared for transfection. The remaining 2 mL of HeLa cells are passaged into another set of three 10 mL lab plates. 660 µL of HeLa cells are passaged into each 10 mL lab plate.

3.4 TRANSFECTION

Two transfection trials were performed. Each transfection occurred when the cell confluency was 80% or above. The first transfection trial was performed on HeLa cells cultured

on 10 mL lab plates. The second transfection trial was performed on HeLa cells cultured on 6-well lab plates. Seven groups of HeLa cells were prepared for transfection. Each group was transfected with different combinations of recombinant plasmids. Group 1 to group 4 were control groups that were transfected with only one of the recombinant plasmids. Group 1, group 2, group 3, and group 4 were transfected with plasmid 4, plasmid 1, plasmid 2, and plasmid 3 respectively. Group 5 to group 7 were the experimental groups that followed the transfection design from **Figure 4**. Group 5 was transfected with plasmid 1 and plasmid 4. Group 6 was transfected with plasmid 2 and plasmid 4. Group 7 was transfected with plasmid 3 and plasmid 4.

Table 1. Transfection Trial 1 of Designed Recombinant Plasmids

Group Number	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
Recombinant Plasmid Number	Plasmid 4	Plasmid 1	Plasmid 2	Plasmid 3	Plasmids 1 & 4	Plasmids 2 & 4	Plasmids 3 & 4
Recombinant Plasmid Amount (μg)	3	6	6	6	3 + 6	3 + 6	3 + 6
Recombinant Plasmid Volume (μL)	21.47	95.39	140.19	92.17	21.47 + 95.39	21.47 + 140.19	21.47 + 92.17
Lipofectamine 2000 Volume (μL)	9	18	18	18	27	27	27

Table 1. Transfection trial 1 of designed recombinant plasmids. The ratio of recombinant plasmid to Lipofectamine 2000 was 1:3. This transfection was performed on HeLa cells cultured on 10 mL lab plates. Seven groups of HeLa cells were prepared. Group 1, group 2, group 3, and group 4 were the control designs. Group 5, group 6, and group 7 were the experimental designs.

Transfection trial 1 was performed with a 1:3 ratio of recombinant plasmid to Lipofectamine 2000 (catalog #11668027) from Thermo Fisher Scientific. While, transfection trial 2 was performed with a 2:5 ratio of recombinant plasmid to Lipofectamine 2000. **Table 1** and **Table 2** list the quantity of reagents used for transfection. Each recombinant plasmid was diluted in 50 μL of Opti-MEM™ I Reduced Serum Medium (catalog #31985062) from Thermo Fisher Scientific. Lipofectamine 2000 was mixed in 50 μL of Opti-MEM™ Medium and

incubated at room temperature for 5 minutes. The diluted recombinant plasmid and diluted Lipofectamine 2000 were combined. This mixture was incubated for 20 minutes at room temperature. Finally, each mixture was applied to the designated group of HeLa cells. The transfected HeLa cells were placed in a 37°C incubator and monitored for protein expression.

Table 2. Transfection Trial 2 of Designed Recombinant Plasmids

Group Number	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
Recombinant Plasmid Number	Plasmid 4	Plasmid 1	Plasmid 2	Plasmid 3	Plasmids 1 & 4	Plasmid 2 & 4	Plasmids 3 & 4
Recombinant Plasmid Amount (μg)	4	4	4	4	2 + 2	2 + 2	2 + 2
Recombinant Plasmid Volume (μL)	114.61	93.02	89.89	112.68	57.31 + 46.51	57.31 + 44.94	57.31 + 56.34
Lipofectamine 2000 Volume (μL)	10	10	10	10	10	10	10

Table 2. Transfection trial 2 of designed recombinant plasmids. The ratio of recombinant plasmid to Lipofectamine 2000 was 2:5. This transfection was performed on HeLa cells cultured on 6-well lab plates. Seven groups of HeLa cells were prepared. Group 1, group 2, group 3, and group 4 were the control designs. Group 5, group 6, and group 7 were the experimental designs.

3.5 IDENTIFYING EXPRESSION

After transfection, the HeLa cells were monitored for GFP expression. The HeLa cells were observed under green fluorescent light. If fluorescence was detected from the HeLa cells, then this indicates that the LPETG motif and Gly5 chain interacted with each other in the presence of Sortase A. This interaction means that the N-terminus and C-terminus of GFP ligated back together to express the protein [45].

HeLa cells were observed every 12 hours after transfection for fluorescence. When checking the HeLa cells for fluorescence, images of HeLa cells were taken on a fluorescence microscope. For transfection trial 1, fluorescence images were taken for 72 hours. While fluorescence images were taken for 48 hours in transfection trial 2. The fluorescence microscope

was set to phase mode and GFP mode to identify protein expression. All fluorescence imaging parameters of phase mode and GFP mode were selected manually. For transfection trial 1, the phase mode was set to an exposure time of 7.579 ms and a gain of 8.7 dB. The GFP mode was set to an exposure time of 334.4 ms and a gain of 14.5 dB. For transfection trial 2, the phase mode was set to an exposure time of 16.16 ms and a gain of 2 dB. The GFP mode was set to an exposure time of 1 s and a gain of 4.4 dB.

CHAPTER 4 - DATA ANALYSIS

HeLa cells were transfected according to **Table 1** and **Table 2**. Transfected HeLa cells were observed every 12 hours. Reconstruction of whole GFP was indicated by GFP fluorescence. If GFP fluorescence was detected, then *in vivo* protein ligation occurred. During transfection trial 1 (**Table 1**), HeLa cells were observed for 72 hours. HeLa cells in transfection trial 2 (**Table 2**) were monitored for 48 hours. After every 12 hours, images of the transfected HeLa cells were taken under phase mode and GFP mode with a fluorescence microscope. Phase mode includes visible light and GFP mode includes UV light. Fluorescence was seen in HeLa cells that were co-transfected with either plasmid 1 and plasmid 4 or plasmid 2 and plasmid 4.

4.1 TRANSFECTION TRIAL 1: 12 HOURS

Phase mode and GFP mode images in **Figure 9** were taken 12 hours after transfection trial 1. Group1, group 2, group 3, group 4, group 5, group 6, and group 7 show transfected HeLa cells under a fluorescence microscope. The amount of transfection reagent used in the first transfection trial is listed in **Table 1**. Since this was the first time observing the HeLa cells for protein expression of GFP, the parameters for the exposure time of phase mode and GFP mode were unknown. Because of this, these images appear dark due to inexperience with the fluorescence microscope. The images labeled phase were applied with a phase filter and viewed under visible light. The images labeled GFP were applied with a GFP filter and viewed under UV light. Group 1, group 2, group 3, and group 4 are the control designs with transfection of only one recombinant plasmid. Therefore, there should be no fluorescence under GFP mode. Group 5, Group 6, and group 7 are the experimental designs with co-transfection of two recombinant plasmids. These HeLa cells in GFP mode were first excited with UV light and fluorescence emission was detected. Due to this, these groups should show some fluorescence under GFP. Based on these images, it is hard to determine if GFP was expressed, because the images are too dark. Overall, this data is insufficient to state if protein ligation occurred. The HeLa cells need to be observed under higher exposure times for the phase mode and GFP mode in order to record better images.

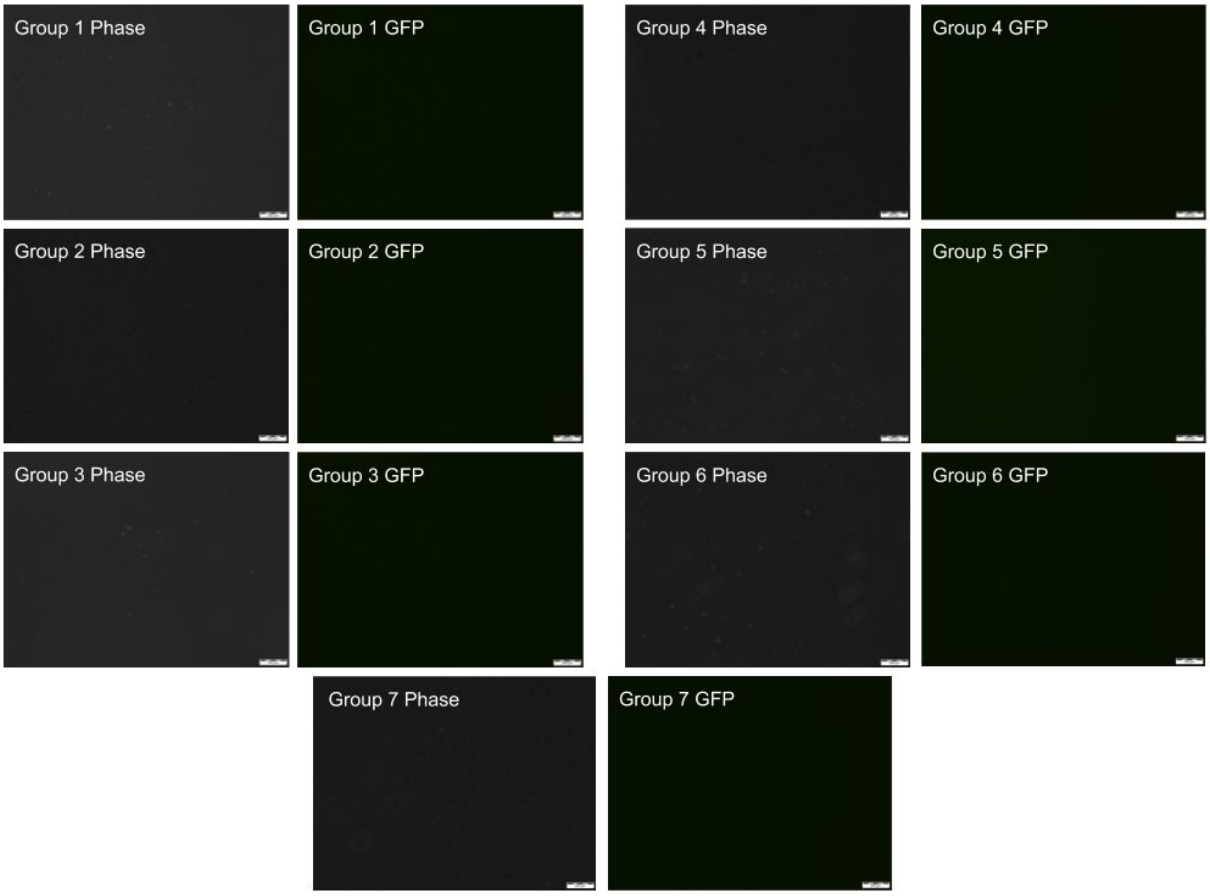


Figure 9. Fluorescence images 12 hour after transfection trial 1. Phase mode and GFP mode were recorded for group 1, group 2 group 3 and group 4, group 5, group 6, and group 7. Group 1, group 2 group 3 and group 4 are the control design, since they are transfected with only one recombinant plasmid. There should be no fluorescence of GFP in these groups. Group 5, group 6, and group 7 are the experimental design, since they are co-transfected with two recombinant plasmids. These groups should show GFP fluorescence. However, these images appeared dark due to low exposure times for phase mode and GFP mode. Based on this data, it is implausible to detect protein expression.

4.2 TRANSFECTION TRIAL 1: 24 HOURS

Fluorescence images were taken under phase mode and GFP mode for group 1, group 2, group 3, group 4, group 5, group 6, and group 7. Group 1, group 2, group 3, and group 4 are transfected with only one recombinant plasmid such as plasmid 4, plasmid 1, plasmid 2, and plasmid 3 respectively. Since these groups are transfected with only one recombinant plasmid, fluorescence of GFP will not be seen. Therefore, these groups are the control designs. Group 5, group 6, and group 7 are co-transfected with two recombinant plasmids. These groups are transfected with the following combinations of recombinant plasmids in this respective order: plasmid 1 & plasmid 4, plasmid 2 & plasmid 4, plasmid 3 & plasmid 4. Group 5, group 6, and group 7 are the experimental designs, so fluorescence of GFP will be seen. **Figure 10** shows the fluorescence images taken 24 hours after transfection. Although the images are more visible, the fluorescence microscope is not properly focused on the HeLa cells. Many of the dots and circles

in the images represent the remaining Lipofectamine 2000 reagent that was not transfected or dead HeLa cell debris. There is indication of some HeLa cells in the clustered areas of group 2, group 4, group 5, group 6, and group 7. However, the resolution needs to be increased to focus on the healthy HeLa cells. There are some marks of bright green fluorescence especially in the GFP image of group 6. These bright green spots are not indications of protein expression. Instead, these bright green spots are mostly likely dead HeLa cell debris. Since the fluorescence microscope is out of focus, it could be picking up dead HeLa cells debris that floated to the top of the cell culture. This data is not concrete enough to support that there is an expression of GFP. In order to improve the results, the fluorescence microscope needs to be adjusted to the correct resolution to view the healthy and adherent HeLa cells on the bottom of the cell culture.

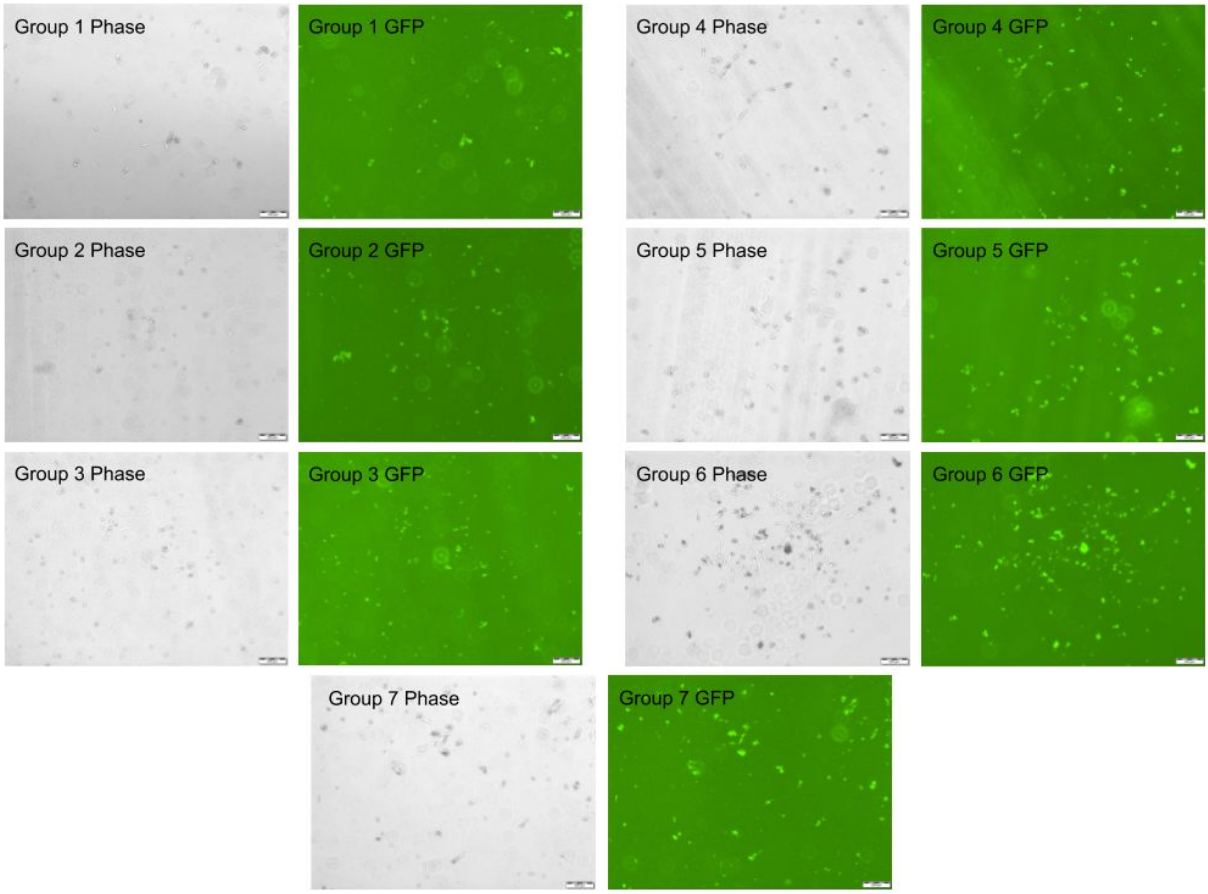


Figure 10. Fluorescence images 24 hours after transfection trial 1. Group 1, group 2, group 3, and group 4 were transfected with plasmid 4, plasmid 1, plasmid 2, and plasmid 3 respectively. Group 5, group 6, and group 7 are co-transfected with the following combinations of plasmids respectively: plasmid 1 & plasmid 4, plasmid 2 & plasmid 4, and plasmid 3 & plasmid 4. The HeLa cells in these images are out of focus. The circles and dots shown in the GFP and phase images represent the untransfected Lipofectamine 2000 reagent. The bright green spots in clustered areas are most likely dead HeLa cell debris that floated to the top of the cell culture. There is not enough evidence to identify GFP fluorescence. In order to determine protein expression, the resolution of the fluorescence microscope needs to increase to capture the adherent HeLa cells.

4.3 TRANSFECTION TRIAL 1: 36 HOURS

After 36 hours of transfection, the HeLa cells were observed for GFP expression under a fluorescence microscope. The fluorescence microscope was set to a higher resolution to capture adherent HeLa cells. These HeLa cells were photographed in phase mode and GFP mode shown in **Figure 11**. The fluorescence images display a better visualization of the HeLa cells, however it is still hard to indicate any expression of GFP in group 5, group 6, and group 7. There are two main reasons for the lack of fluorescence detection. First, the confluency of the cell culture is low. This could be caused from transfecting the HeLa cells in 10 mL lab plates. 10 mL lab plates have a big surface area that is not ideal for observing fluorescence, because HeLa cells are more scattered apart from each other. Due to this, the Lipofectamine 2000 was not able to transfect the

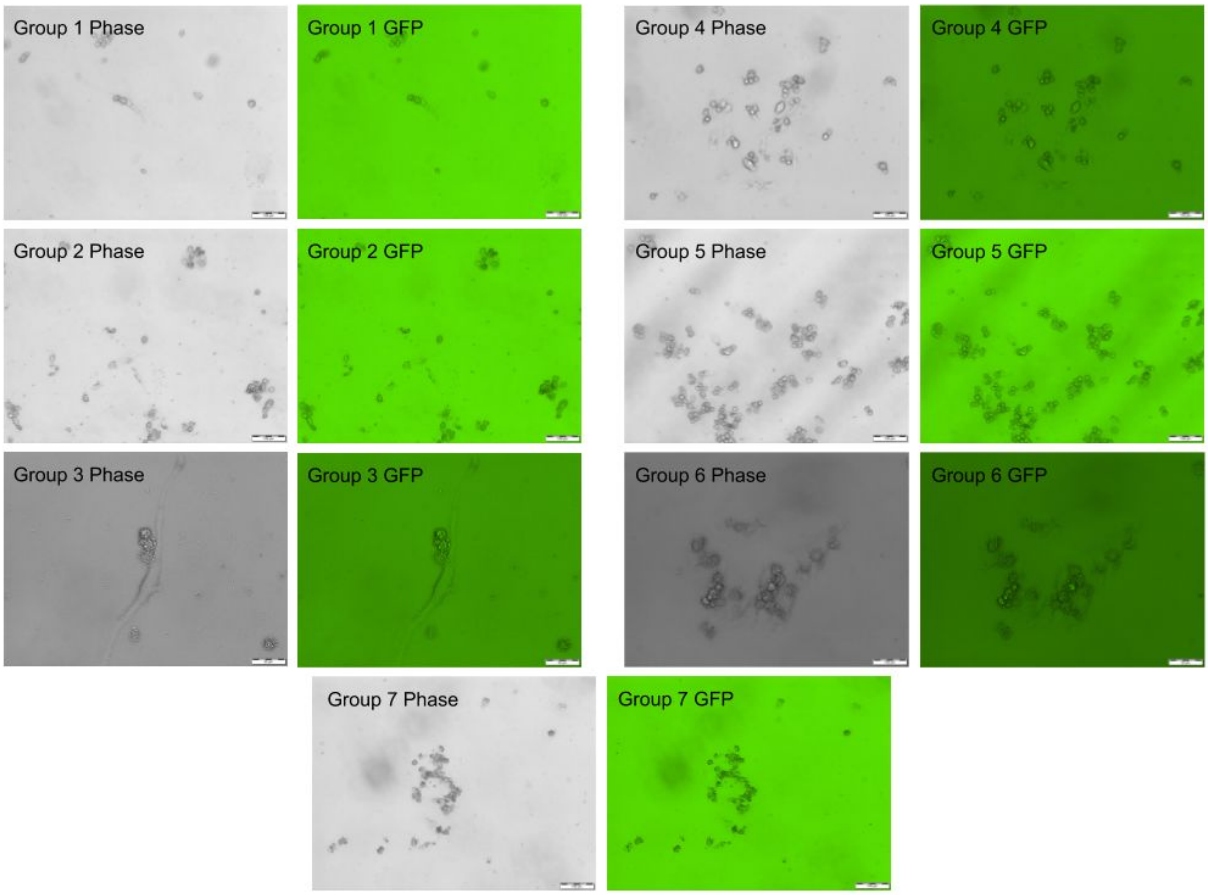


Figure 11. Fluorescence images 36 hours after transfection trial 1. Fluorescence of GFP is still not detected in group 5, group 6, and group 7. Transfection trial 1 was performed on HeLa cells cultured on 10 mL lab plates. Therefore, the HeLa cells are scattered apart on a bigger surface area and the Lipofectamine 2000 did not transfect the recombinant plasmid throughout all the HeLa cells. Also, the fluorescence microscope was set to automatic when capturing images of phase mode and GFP mode. Due to this, the background appears too bright to observe any GFP fluorescence. The fluorescence microscope needs to be set to manual to adjust the exposure times. Expression of GFP can not be determined from this data.

recombinant plasmid throughout all the HeLa cells. HeLa cells in specific areas of the 10 mL lab plate will be transfected successfully, while HeLa cells in other areas will not be transfected. Also, it is not advisable to increase the amount of transfection reagent, because it is toxic and can increase the chances of killing HeLa cells. Therefore, it is better to transfect HeLa cells in a small surface area with a limited amount of transfection reagent. This way, the HeLa cells are closer and packed together. The Lipofectamine 2000 can transfect the recombinant plasmid evenly across all the HeLa cells in a smaller region. Then there will be a better chance of seeing fluorescence. Second, the fluorescence microscope was set to automatic, thus the images appear brighter. The phase mode and GFP mode have a longer exposure time resulting in brighter images. Although the resolution of the fluorescence microscope is high enough to focus on the adherent HeLa cells, the background is too light to identify any fluorescence. The fluorescence microscope needs to be set to manual when recording the phase mode and GFP mode. Then, the exposure times can be controlled to prevent a bright background and the fluorescence will be more distinguishable. Overall, **Figure 11** does not indicate expression of GFP. GFP mode images of group 5, group 6, and group 7 do not show any fluorescence.

4.4 TRANSFECTION TRIAL 1: 48 HOURS

During imaging the HeLa cell under phase mode and GFP mode for the fourth time, the fluorescence microscope was set to a higher resolution to focus on the adherent HeLa cells on the bottom of the cell culture. Also, the exposure times were manually adjusted to control the brightness of the background. For phase mode, the exposure time was 7.579 ms and the gain was 8.7 dB. For GFP mode, the exposure time was 334.4 ms and the gain was 14.5 dB. Based on these parameters, fluorescence is seen on the GFP images of group 5 and group 6. Group 1, group 2, group 3, group 4, and group 7 do not show much adherent HeLa cells, because the majority of the HeLa cells have died after 48 hours. The remaining HeLa cells are in the early stages of mutation. This is seen in group 5 and group 6 of **Figure 12**. The phase mode of group 5 and group 6, show some deformed HeLa cells. These deformed HeLa cells are an older generation that were transfected successfully. The circled areas in red show some green fluorescence. The green fluorescence indicates that GFP was expressed. This makes sense, since group 5 was co-transfected with plasmid 1 and plasmid 4. This was the original experimental design. While group 6 was co-transfected with plasmid 2 and plasmid 4 that was the modified design with TOP1 and SH3 domains. These two domains have a high binding affinity with each other. The expression of GFP infers that *in vivo* protein ligation occurred and fluorescence is visible by 48 hours.

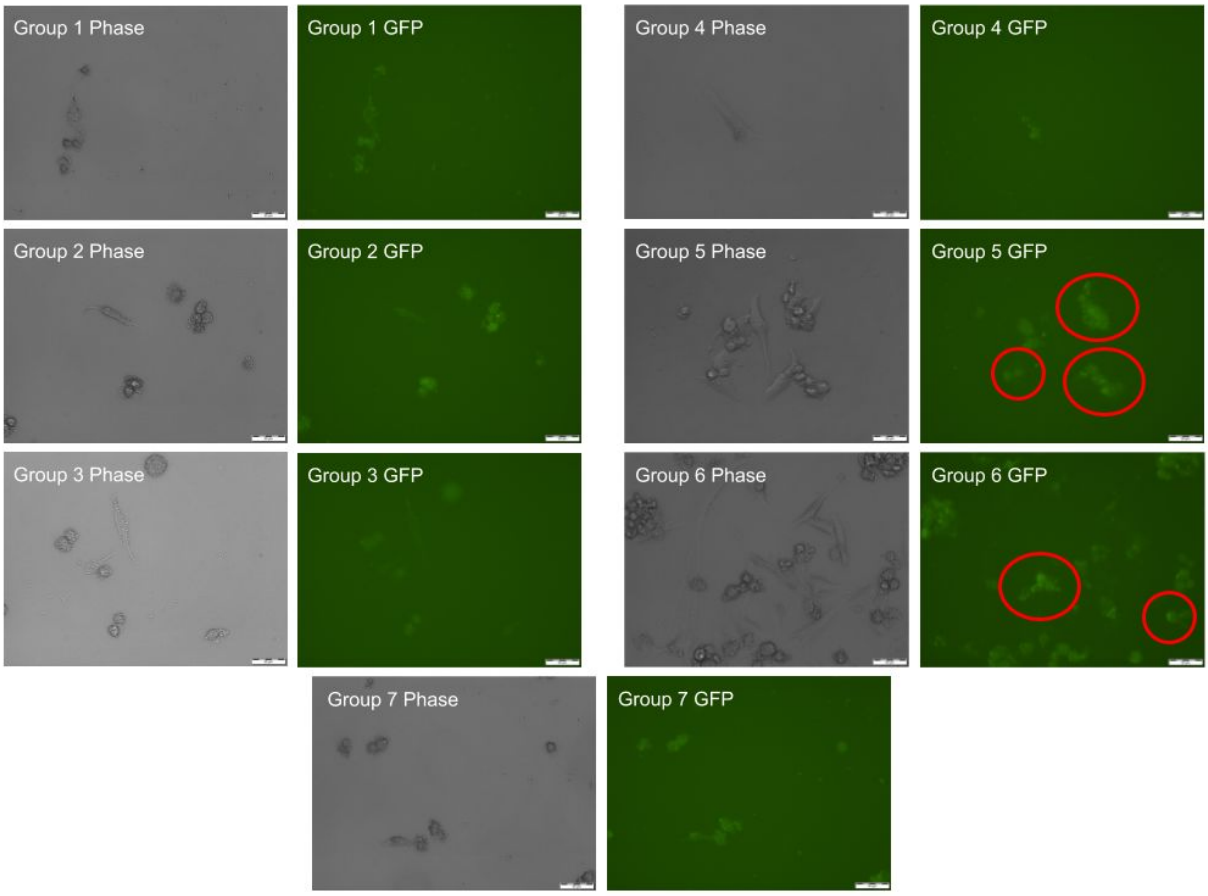


Figure 12. Fluorescence images 48 hours after transfection trial 1. The fluorescence microscope was manually set to an exposure time of 7.579 ms and a gain of 8.7 dB for phase mode. For GFP mode, the exposure time is 334.4 ms and a gain of 14.5 dB. After 48 hours, the majority of the HeLa cell in group 1, group 2, group 3, group 4, and group 7 died. The remaining HeLa cells are in the early stages of mutation. Group 5 and group 6 show green fluorescence. The red circled areas mark fluorescence of GFP. These areas are older generations of HeLa cells that were transfected successfully. This data indicates that protein ligation occured and GFP was expressed.

4.5 TRANSFECTION TRIAL 1: 60 HOURS

The same exposure times and gains were manually set on the fluorescence microscope. The exposure time was 7.579 ms and the gain was 8.7 dB for phase mode. The exposure time was 334.4 ms and the gain was 14.5 dB for GFP mode. By 60 hours after transfection, there were no clusters present in group 1, group 2, group 3, group 4, and group 7. This is visible in **Figure 13**. The phase mode images capture one or two HeLa cells. Group 5 and group 6 have a higher HeLa cell count, however they are more deformed. The phase mode images show the spread of mutated HeLa cells. At this point, the mutation has taken over and there are no healthy HeLa cells. The GFP mode images show a decrease in the green fluorescence. **Figure 13** displays fewer areas with GFP fluorescence that are circled in red. The number of mutated HeLa cells is more abundant in **Figure 13** compared to the HeLa cells present in **Figure 12**. Also, the integrity

and shape of the HeLa cells is completely distorted. The HeLa cells have lost their fibroblast shape. Based on this data, the cut off point of transfection is 48 hours. At 48 hours, the HeLa cells contain the highest level of expression with minimum signs of mutation. Overall, at 60 hours, the HeLa cells decreased the expression of GFP and began to die. Only the mutated HeLa cells survived and grew.

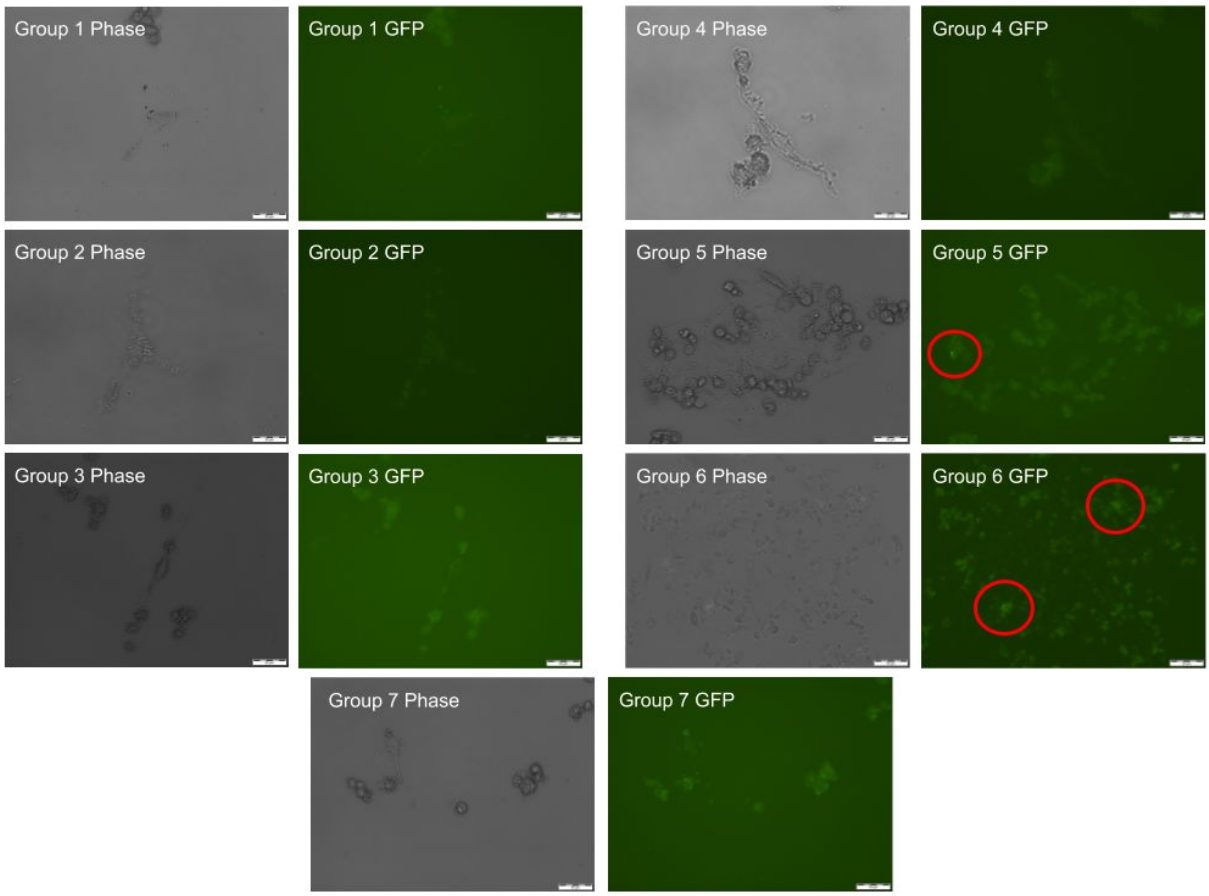


Figure 13. Fluorescence images 60 hours after transfection trial 1. The fluorescence microscope was set to the same parameters when recording phase mode and GFP mode. The adherent HeLa cells have mutated and began to spread. Group 1, group 2, group 3, group 4, and group 7 contain individual deformed HeLa cells. Group 5 and group 6 display clusters of HeLa cells that lost the fibroblast shape. The GFP fluorescence is circled in red and it appears the HeLa cells are losing the expression of GFP. This data indicates that the HeLa cells are completely mutated by 60 hours.

4.6 TRANSFECTION TRIAL 1: 72 HOURS

After 72 hours of transfection trial 1, there are no more healthy HeLa cells. The cell culture has changed to a dark orange color. The remaining HeLa cells have completely mutated and began to die. **Figure 14** displays the fluorescence images of the mutated HeLa cells. The fluorescence microscope recorded phase mode and GFP mode with the same parameters from 48 hours and 60 hours. The HeLa cell clusters decreased and fluorescence of GFP is gone. Group 1 shows HeLa cell debris left from the mutation. Group 2, group 3, group 4, and group 7 present

isolated mutated HeLa cells that have lost the fibroblast shape. Group 5 shows a small group of mutated HeLa cells. The phase mode image does not show any healthy HeLa cells. The circles and spots in the phase mode images represent the mutations. The GFP mode image does not indicate any fluorescence. Group 6 presents late stages of mutated HeLa cells. The phase mode image contains mutated HeLa cells with some debris. The GFP mode image does not suggest protein expression. By 72 hours, the HeLa cells in each group stopped expressing the recombinant plasmid.

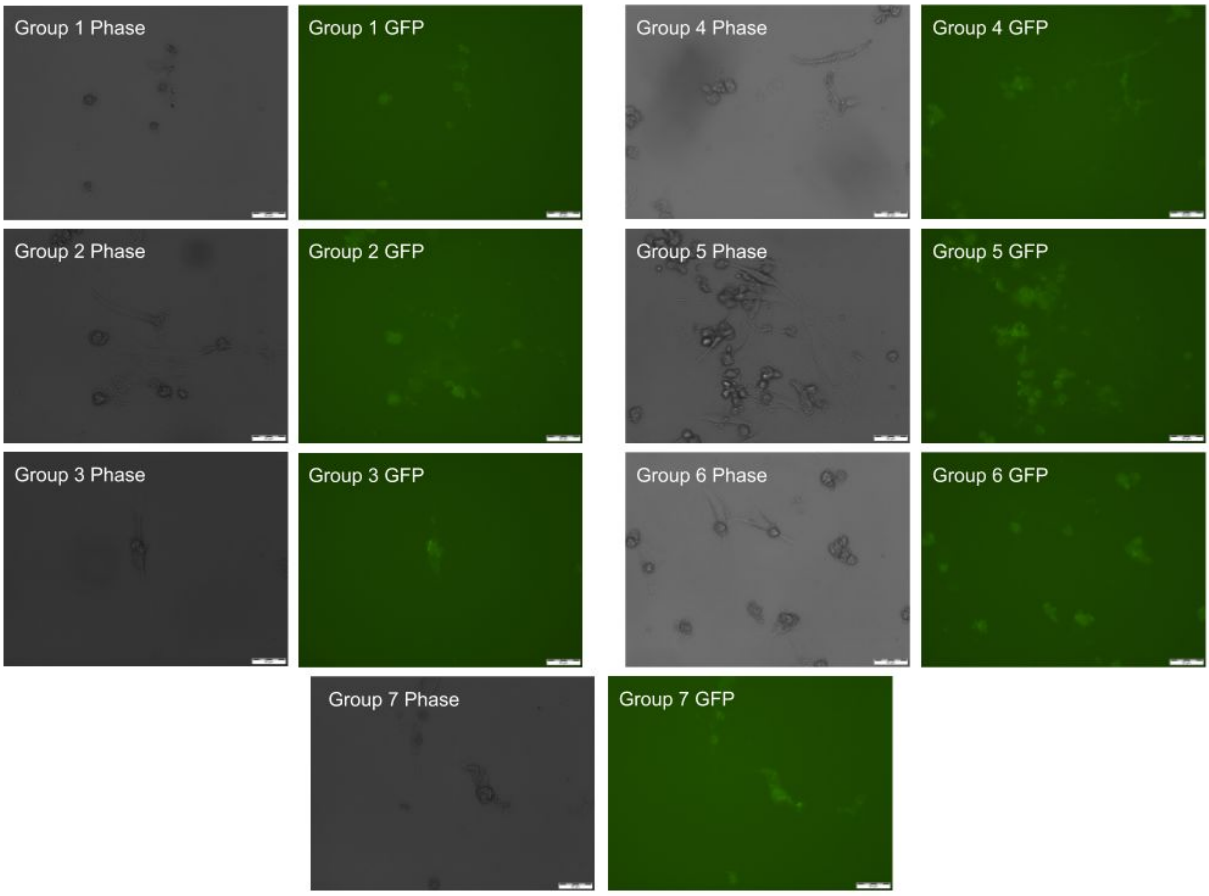


Figure 14. Fluorescence images 72 hours after transfection trial 1. The same parameters were applied to record the HeLa cells in phase mode and GFP mode. Majority of the HeLa cells in each group have died and the remaining HeLa cells are mutated. Group 5, group 6, and group 7 do not contain any expression of GFP. The phase mode images display deformed HeLa cells without a fibroblast shape and dead HeLa cell debris. The GFP mode images do not contain any fluorescence. At this point, the mutation has contaminated the cell culture and prevented protein ligation.

4.7 TRANSFECTION TRIAL 2: 12 HOURS

For the second trial of transfection, the HeLa cells were cultured in two 6-well lab plates as portrayed in **Figure 8**. The HeLa cells were transfected in a smaller surface area. Therefore there is a higher chance the Lipofectamine 2000 will transfect the recombinant plasmid throughout all the HeLa cells. The amount of transfection reagent used in the second transfection

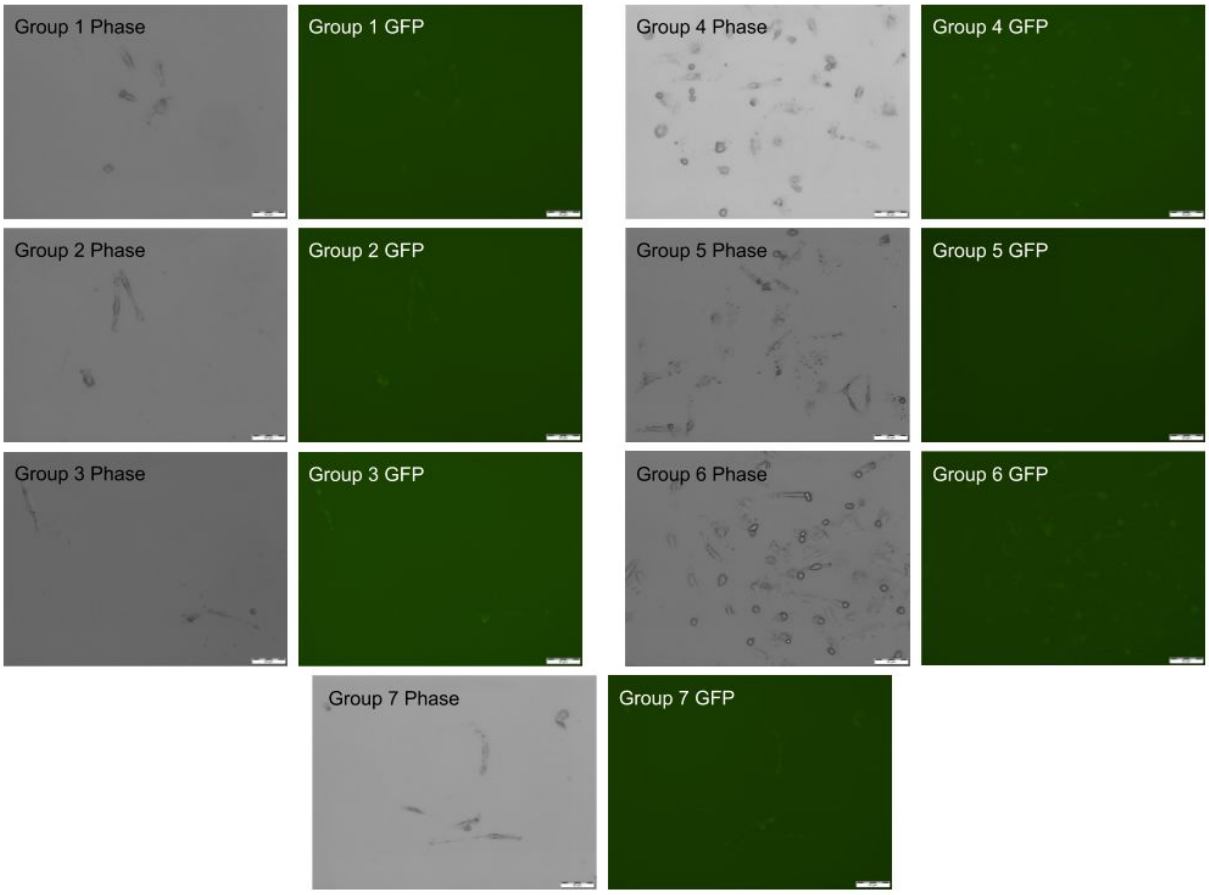


Figure 15. Fluorescence images 12 hours after transfection trial 2. HeLa cells were transfected on two 6-well lab plates. Fluorescence images were taken of phase mode and GFP mode. The exposure time was 16.16 ms and gain was 2 dB for phase mode. The exposure time was 1 s and gain was 4.4 dB for GFP mode. Group 1, group 2, group 3, and group 7 display a low HeLa cell count, because the cell culture media was replaced with fresh media after 12 hours. This interrupted the transfection process, by removing the Lipofectamine 2000 and Opti-MEM™ Medium from the cell culture. Group 4, group 5, and group 6 have a higher cell count. However, there is no visible fluorescence to indicate expression of GFP. This data is inconclusive to determine if protein ligation occured. The HeLa cells need to be observed again after more time for protein expression.

trial is listed in **Table 2**. After 12 hours of transfection trial 2, fluorescence images were taken. The fluorescence microscope was set to manual from the start. **Figure 15** displayed the phase mode and GFP mode images of HeLa cells from each group. The exposure time was 16.16 ms and the gain was 2 dB for phase mode. The exposure time was 1 s and the gain was 4.4 dB for GFP mode. The cell culture media was replaced with fresh media after 12 hours. This creates an issue, since the excess Lipofectamine 2000 and Opti-MEM™ Medium were discarded. The HeLa cells need to sit in the transfection reagent for at least 24 hours for the most optimal results. Since the cell culture media with the Lipofectamine 2000 and Opti-MEM™ Medium was replaced too early, the transfection procedure was interrupted. This explains the reason for the low HeLa cell count in the phase mode images of group 1, group 2, group 3, and group 7. Group 4, group 5, and group 6 maintain a higher HeLa cell count. Group 1, group 2, group 3, and group

4 were the control designs, so there is no fluorescence seen in the GFP mode images. Groups 5, group 6, and group 7 were the experimental designs, so there should be fluorescence of GFP. However, because the transfection process was interrupted, the green fluorescence might appear more faint. **Figure 15** does not portray any GFP fluorescence. Thus, the data indicate that there is no expression of GFP and protein ligation did not occur. The HeLa cells require more time to express GFP.

4.8 TRANSFECTION TRIAL 2: 24 HOURS

After 24 hours of transfection, there is faint fluorescence visible. The fluorescence microscope was adjusted to the same parameters from 12 hours to capture images in phase mode

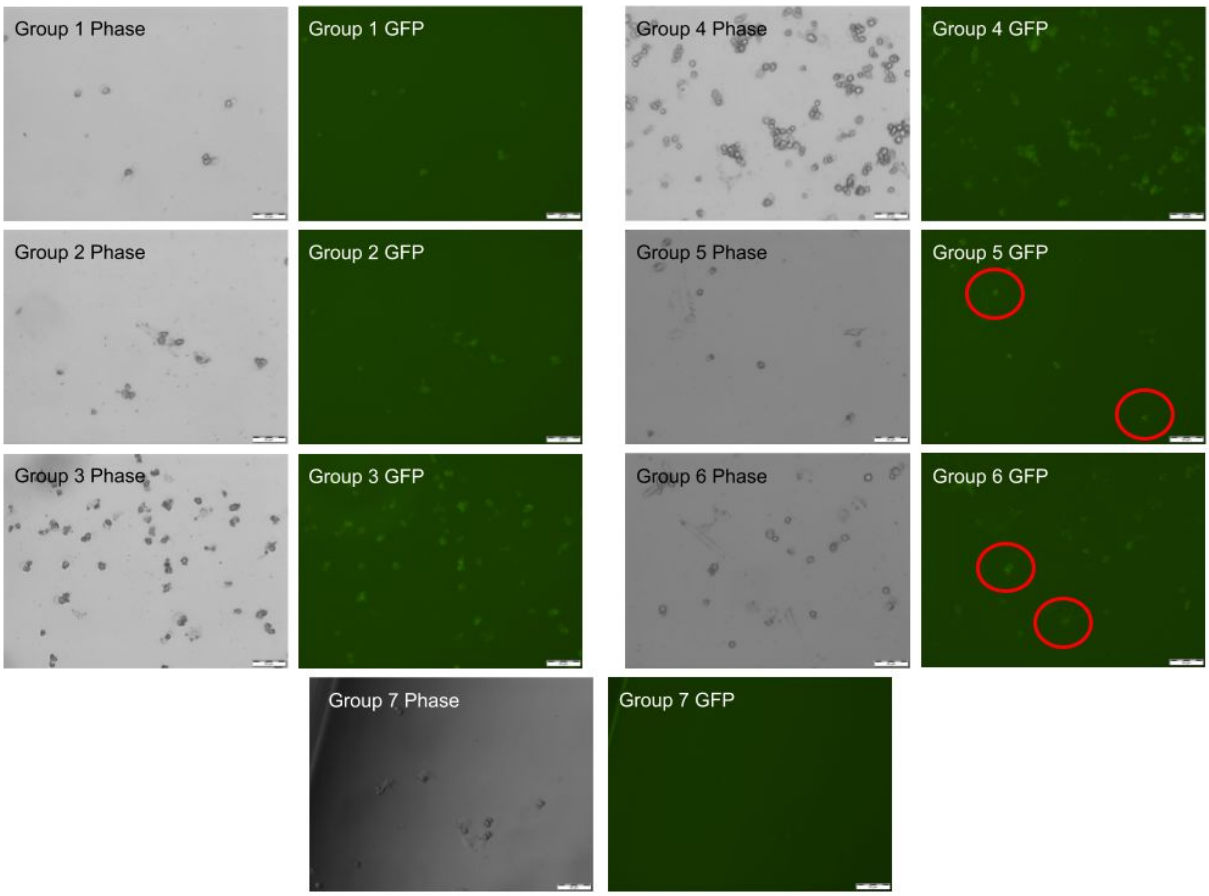


Figure 16. Fluorescence images 24 hours after transfection trial 2. The HeLa cells started to grow and form small colonies. Each group contains healthy HeLa cells with a fibroblast shape. The control groups do not indicate any green fluorescence. This is expected, because the control groups were not transfected with the recombinant plasmid encoding for Sortase A. Therefore, protein ligation will not be activated to express GFP. The experimental groups portray faint fluorescence. The reason for this is due to an interruption in the transfection procedure. The cell culture with Lipofectamine 2000 and Opti-MEM™ Medium was discarded after 12 hour of transfection. Thus, the transfection reagent did not transfer the recombinant plasmids to all the HeLa cells. Only a few HeLa cells were transfected successfully and able to express GFP. The GFP mode images of group 5 and group 6 circle the areas of protein expression. The fluorescence microscope was set to the same parameters from 12 hours to capture the phase mode and GFP mode of the HeLa cells.

and GFP mode. Over time, the HeLa cells started to grow again and colonies are present in group 3, group 4, group 5, and group 6. The HeLa cells still maintain a fibroblast structure in each group. The control groups do not show any green fluorescence. This is expected, because the control groups were transfected with only one recombinant plasmid. There is no presence of Sortase A in group 1, group 2, group 3, and group 4, so protein ligation did not happen. The expression of GFP did not take place in the control groups. There is faint fluorescence in the experimental groups. **Figure 16** circles in red the areas of protein expression in group 5 and group 6. The green fluorescence is not as intense from the previous transfection trial, because the cell culture media was replaced too early. The cell culture media was replaced after 12 hours, so the Lipofectamine 2000 and Opti-MEM™ Medium were discarded. The transfection process was interrupted halfway, resulting in a low number of recombinant plasmids transferred into the HeLa cells. Group 7 is part of the experimental design with BOT1 and SH3 domains. These two domains have no binding affinity, so there is a low chance of protein ligation taking place. This is depicted in **Figure 16**, since the GFP mode images do not show any green fluorescence. The data specifies that only a small number of HeLa cells were transfected successfully and able to express GFP. The HeLa cells need to sit in the transfection reagent for at least 24 hours to obtain a higher level of protein expression.

4.9 TRANSFECTION TRIAL 2: 36 HOURS

Figure 17 displays the fluorescence images of the HeLa cells from each group. The fluorescence microscope took images under phase mode and GFP mode of each group. The exposure times and gains for phase mode and GFP mode were adjusted manually. These parameters are the same as those images taken in 12 and 24 hours. After 36 hours of transfection, the HeLa cells begin to mutate and die. The phase mode images of all the groups display a low HeLa cell count. The GFP mode images of group 1, group 2, group 3, and group 4 do not contain any green fluorescence, because they are the control design. Among the experimental groups, group 5 shows fluorescence of GFP circled in red. The intensity of the green fluorescence is higher in **Figure 17** compared to the green fluorescence in **Figure 16**. Group 6 and group 7 do not indicate any fluorescence of GFP. Overall, protein expression is only present in group 5. The data infers that the best time to change the cell culture media with fresh media is after 24 hours. This way, the transfection procedure will not be interrupted. The Lipofectamine 2000 and Opti-MEM™ Medium will transfer the recombinant plasmids to the majority of HeLa cells. Also, the HeLa cells will continue to grow without mutating and fluorescence of GFP can be detected in group 5, group 6, and group 7, because these are the experimental design.

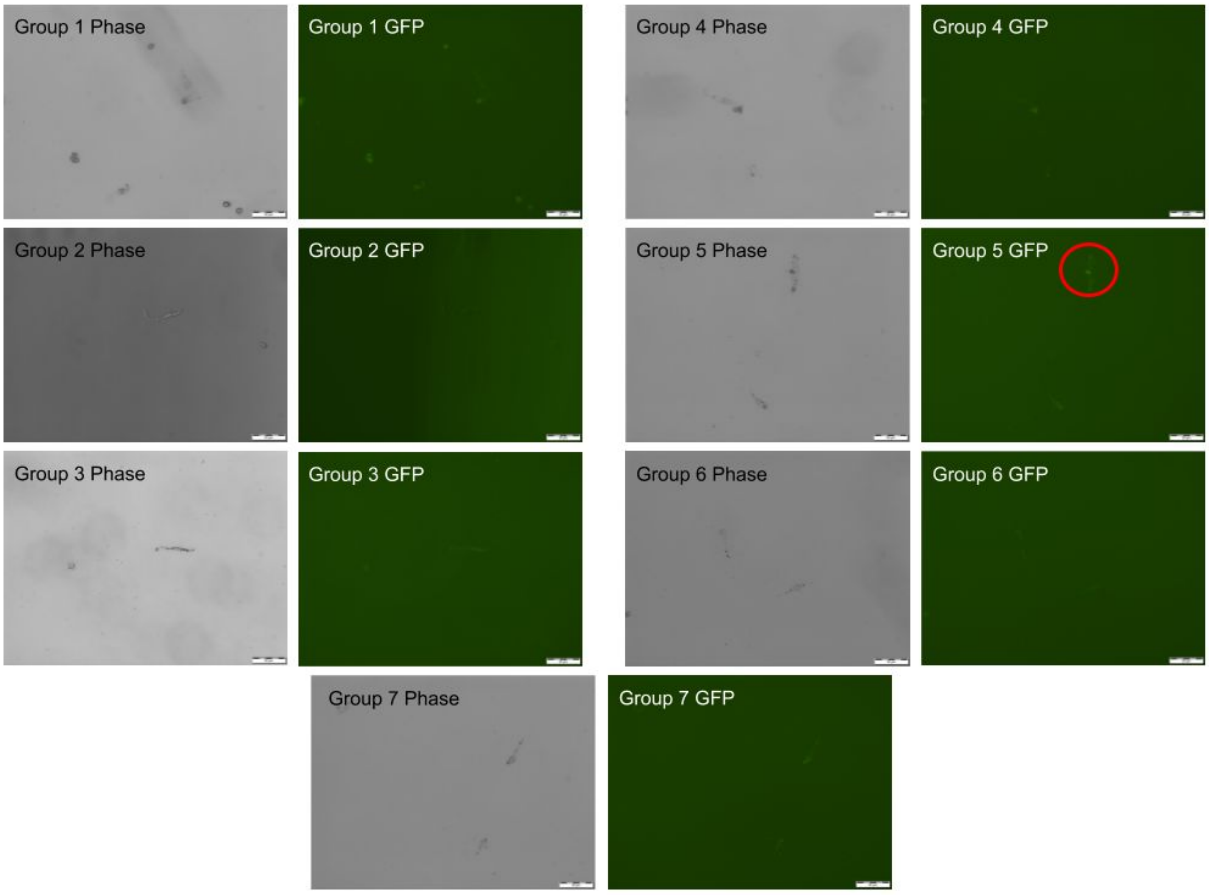


Figure 17. Fluorescence images 36 hours after transfection trial 2. The fluorescence microscope was set to the same manual setting from 12 and 24 hours to create these phase mode and GFP mode images. The HeLa cell count in all the groups have decreased. Only group 5 contains fluorescence of GFP circled in red. This green fluorescence is more intense compared to the green fluorescence after 24 hours.The rest of the experimental groups do not display protein expression, because the transfection procedure was interrupted. The cell culture media with Lipofectamine 2000 and Opti-MEM™ Medium was replaced with fresh media too early. This data specifies that the recombinant plasmids were not transfected into all the HeLa cells from each group. Protein ligation only occured in group 5, so it was able to express GFP.

4.10 TRANSFECTION TRIAL 2: 48 HOURS

After 48 hours of transfection, the green fluorescence in group 5 became more intense. **Figure 18** circles the areas of protein expression in red. The same parameters were applied to create the phase mode and GFP mode images. The exposure times and gains were consistent from 12, 24, 36, and 48 hours. Group 1, group 2, group 3, group 4, and group 7 have a low HeLa cell count. The reason for this is that the cell culture media with the Lipofectamine 2000 and Opti-MEM™ Medium was changed with fresh media too early. While replacing the cell culture media, the majority of the transfection reagents and HeLa cells were discarded. Because of this, the transfection was not uniform across all the HeLa cells. Majority of the HeLa cells were not transfected properly. The few HeLa cells that were transfected successfully are able to express

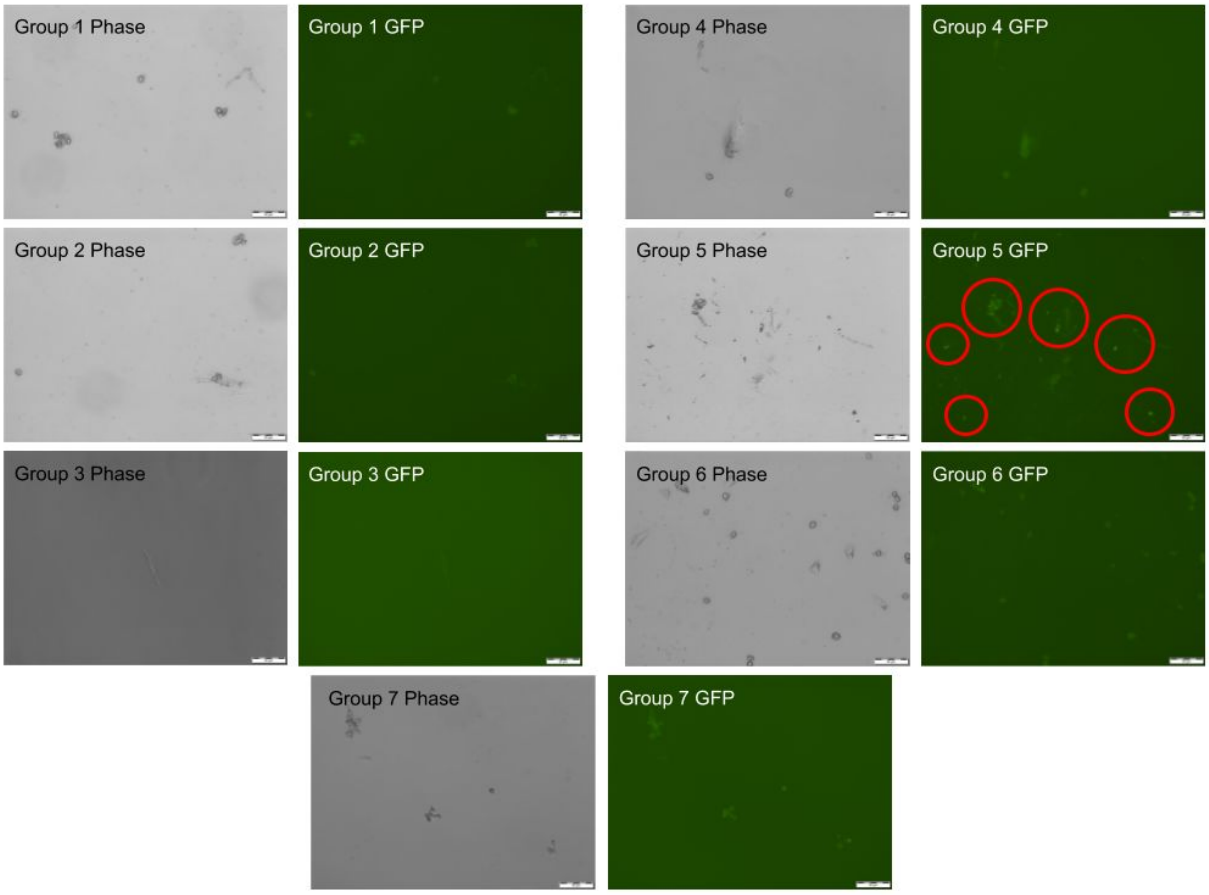


Figure 18. Fluorescence images 48 hours after transfection trial 2. The data support that there is expression of GFP in group 5. The areas of protein expression are circled in red. The images of phase mode and GFP mode were taken under the same conditions as the fluorescence images from 12, 24, and 36 hours Group 1, group 2, group 3, group 4, and group 7 have a low HeLa cell count due to changing the cell culture media too early. The Lipofectamine 2000 and Opti-MEM™ Medium in the cell culture was discarded that interrupted the transfection process. The recombinant plasmid was successfully transfected in a few HeLa cells. This explains the absence of protein expression in group 6 and group 7. There is strong evidence that the transfection was successful in only group 5, because the green fluorescence is more intense compared to the images in the previous transfection trial. This claims that plasmid 1 and plasmid 4 were co-transfected and Sortase A was expressed to activate in-vivo protein ligation between the LPETG motif and Gly5 chain.

GFP. Group 5 and group 6 have a higher HeLa cell count. The HeLa cells in these groups were able to grow and form small colonies. The HeLa cells do not show any signs of mutation. Group 5 is the only set of HeLa cells that show fluorescence of GFP. The transfected HeLa cells have divided and protein expression is seen in multiple areas. The data state that *in vivo* protein ligation occurred, because the green fluorescence is more intense. The GFP mode image of group 5 displays a very evident green fluorescence compared to the images from the previous transfection trial. The data implies that the co-transfection with plasmid 1 and plasmid 4 worked and Sortase A was expressed to activate *in vivo* protein ligation between the LPETG motif and Gly5 chain. This interaction allowed the N-terminus and C-terminus of GFP to bind together.

Group 6 and group 7 were expected to show GFP fluorescence. However, there is no green fluorescence. This could be due to the interruption in the transfection process. The HeLa cells in group 6 and group 7 were not co-transfected properly that led to an absence of fluorescence. Overall, there is strong evidence that protein expression is present in only group 5. **Figure 18** portrays a more intense green fluorescence that supports the claim that *in vivo* protein ligation was initiated and GFP was expressed.

4.11 RESULTS

Table 3 summarizes the GFP fluorescence results from each transfection trial. The sections highlighted in red do not show fluorescence, while sections highlighted in green indicate fluorescence of GFP. Groups 1, group 2, group 3, and group 4 are the control designs that do not express GFP. This is expected, because Sortase A is not introduced in these groups. The data proves this throughout each transfection trial. Groups 5, group 6, and group 7 are the experimental designs that are co-transfected with one of the three designed recombinant plasmids and Sortase A. Group 5 is the original design with the LPETG motif and Gly5 chain. In the presence of Sortase A, bioconjugation will occur. The N-terminus and C-terminus of GFP will ligate together to afford the whole protein. Group 6 contains the TOP1 and SH3 domains that will increase specificity and efficiency of ligation, because of selective binding [38]. In contrast, group 7 contains the BOT1 and SH3 domains that are unable to reconstruct whole GFP due to no binding affinity [39].

Table 3. Summary of GFP Fluorescence Results

Group Number	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
Transfection Trial 1: 12 Hours	No GFP	No GFP	No GFP	No GFP	No GFP	No GFP	No GFP
Transfection Trial 1: 24 Hours	No GFP	No GFP	No GFP	No GFP	No GFP	No GFP	No GFP
Transfection Trial 1: 36 Hours	No GFP	No GFP	No GFP	No GFP	No GFP	No GFP	No GFP
Transfection Trial 1: 48 Hours	No GFP	No GFP	No GFP	No GFP	GFP	GFP	No GFP

Transfection Trial 1: 60 Hours	No GFP	No GFP	No GFP	No GFP	GFP	GFP	No GFP
Transfection Trial 1: 72 Hours	No GFP	No GFP	No GFP	No GFP	No GFP	No GFP	No GFP
Transfection Trial 2: 12 Hours	No GFP	No GFP	No GFP	No GFP	No GFP	No GFP	No GFP
Transfection Trial 2: 24 Hours	No GFP	No GFP	No GFP	No GFP	GFP	GFP	No GFP
Transfection Trial 2: 36 Hours	No GFP	No GFP	No GFP	No GFP	GFP	No GFP	No GFP
Transfection Trial 2: 48 Hours	No GFP	No GFP	No GFP	No GFP	GFP	No GFP	No GFP

Table 3. Summary of GFP fluorescence results. Sections highlighted in red do not express GFP and sections highlighted in green express GFP. Group 1, group 2, group 3, and group 4 are the control design. Fluorescence of GFP is not detected as expected. Group 5, group 6, and group 7 are the experimental design that are expected to show a different level of GFP expression. In transfection trial 1, Group 5 and group 6 detects fluorescence of GFP after 48 and 60 hours. This is expected, since group 5 is the original design with the LPETG motf and Gly5 chain. Group 6 includes TOP1 and SH3 domains that have a binding affinity with each other. Therefore, group 5 and group 6 indicate protein ligation between the N-terminus and C-terminus of GFP. In transfection trial 2, group 5 and group 6 show fluorescence of GFP after 24 hours. However, after 36 and 48 hours, only group 5 shows GFP fluorescence. This is resulted from the interruption in the transfection procedure causing a lack of HeLa cells expressing both recombinant plasmids.

In transfection trial 1 (**Table 1**), GFP fluorescence is seen after 48 and 60 hours in only groups 5 and 6. After 72 hours of transfection, the majority of the HeLa cells have died and the remaining HeLa cells are completely mutated. In transfection trial 2 (**Table 2**), groups 5 and 6 express GFP after 24 hours. However, after 36 and 48 hours only group 5 has fluorescence of GFP. The reason for this is due to an interruption in the transfection procedure. Most of the HeLa cells in group 5, group 6, and group 7 were not co-transfected successfully. The few HeLa cells that were co-transfected successfully, showed fluorescence of GFP like in group 5. The intensity of the fluorescence increases in group 5. This data infers that the co-transfection of the recombinant plasmids in group 5 expressed Sortase A, LPETG motif, and Gly5 chain while performing the necessary post-translational modifications [46]. There is evidence that the substrate specificity of the LPETG motif and Gly5 chain in HeLa cells chain functions the same

way as in gram-positive bacterial cells [47]. The recombinant plasmids were expressed and Sortase A spliced between threonine and glycine on the LPETG motif. The carboxyl group on the threonine cross-links with the Gly5 chain. This bioconjugation allows expression of GFP by bringing the N-terminus and C-terminus of GFP. Finally, expression of GFP is evident from the fluorescence of HeLa cells.

4.12 CONCLUSION

Overall, there is evidence that this design, *in vivo* non-ribosomal protein synthesis in mammalian cells, works. **Figure 3** depicts the schematic of three recombinant plasmids. Each recombinant plasmid includes the GFP sequence spliced into two segments. The original design in plasmid 1 includes the LPETG motif on the N-terminus of the GFP gene while the Gly5 chain is inserted on the C-terminus of the GFP gene. When Sortase A is expressed, this enzyme will cleave in between the threonine and glycine on the LPETG motif and initiate bioconjugation between the Gly5 chains and carboxyl group on the threonine. When this interaction occurs, the N-terminus and C-terminus of GFP will come together, ligate back together, and restore the function of this protein. Plasmids 2 and 3 include additional components that facilitate or prohibit protein ligation of GFP. Plasmid 2 includes the TOP1 and SH3 domains that increase the signal of GFP, because they have a high binding affinity with each other [38]. Plasmid 3 includes the BOT1 and SH3 domains that decrease the signal of GFP, because they have a no binding affinity with each other [39]. Each recombinant plasmid is co-transfected into HeLa cells with a fourth recombinant plasmid that encodes only for the Sortase A enzyme. This experimental design is portrayed in **Figure 4**. Reconstruction of whole GFP is determined by green fluorescence from HeLa cells.

The data shows visible green fluorescence in groups 5 and 6 after 48 hours of transfection for trials 1 and 2. As long as green fluorescence is detected in the experimental groups and not in the control groups, then it indicates that the transfection was successful. When these two factors are observed, then there is strong indication that the N-terminus and C-terminus of GFP ligated back together to afford whole GFP. The data supports this claim, since there is no visible green fluorescence in control group 1, group 2, group 3, and group 4 (**Figure 9**, **Figure 10**, **Figure 11**, **Figure 12**, **Figure 13**, **Figure 14**, **Figure 15**, **Figure 16**, **Figure 17**, and **Figure 18**). Based on these results, the original experimental design in plasmid 1 was able to perform *in vivo* protein ligation and produce GFP. This is proven in experimental group 5 (**Figure 12**, **Figure 13**, **Figure 16**, **Figure 17**, and **Figure 18**). Also, the addition of TOP1 and SH3 domains in plasmid 2 achieved bioconjugation. This is specified in experimental group 6 (**Figure 12**, **Figure 13**, and **Figure 16**). However, there were some technical challenges in the transfection procedure that prevented uniform transfection throughout all the HeLa cells in the cell culture. Due to this, not all HeLa cells were co-transfected successfully. This is the reason, the GFP fluorescence is not detected in group 7 with plasmid 3 that includes the BOT1 and SH3 domains. Normally, BOT1 and SHE3 domains would prevent protein ligation, since they do not bind with

each other. Yet, there should be faint green fluorescence, since Sortase A will still trigger the baseline bioconjugation between the LPETG motif and Gly5 chain. Thus, the level of GFP expression will be low in group 7.

CHAPTER 5 - FUTURE WORK

Examining the transfection results under a fluorescence microscope gives a visual representation of the GFP reconstruction. Fluorescence of GFP is recorded in experimental groups 5 and 6. Reconstruction of GFP is not observed in control groups 1, 2, 3, and 4. Based on these results, it is concluded that Sortase A can catalyze protein ligation *in vivo*. Other *in vitro* tests need to be performed to confirm the reconstruction of whole GFP.

5.1 QUANTITATIVE ANALYSIS BY FLOW CYTOMETRY

There are chances that the control groups can show slight green fluorescence. The N-terminus and C-terminus can still come close enough to express GFP. Also, a visualization under the fluorescence microscope does not give a quantitative result on the number of HeLa cells expressing GFP. Therefore, flow cytometry needs to be performed to determine the total of HeLa cells with fluorescence. This method would help in identifying the efficiency of each combination of recombinant plasmids [48]. HeLa cells co-transfected with plasmid 1 and plasmid 4 would have an average amount of fluorescence. This is the original design with only the LPETG motif and Gly5 chain. HeLa cells co-transfected with plasmid 2 and plasmid 4 are expected to have a very high fluorescence, because this design includes TOP1 and SH3 domains. They have a high binding affinity with each other, so there will be a higher level of protein expression. HeLa cells co-transfected with plasmid 3 and plasmid 4 will have a low amount of fluorescence. This design contains BOT1 and SH3 domains that have no binding affinity with each other. Thus, the HeLa cells are expected to show a low level of GFP expression. In order to confirm the difference in protein expression between the experimental designs, the HeLa cells need to be processed through flow cytometry such as FACS. FACS will be able to analyze the transfected HeLa cells and report an actual level of protein expression [49]. This quantitative result can prove the efficiency of GFP expression for each experimental design. FACS has the ability to separate between healthy and mutated HeLa cells [50]. Based on this, FACS will only consider healthy HeLa cells and reject other debris such as dead HeLa cells and excess transfection reagent.

5.2 GFP IDENTIFICATION BY WESTERN BLOT

Once protein expression is identified visually and quantitatively, a western blot needs to be performed to confirm that the protein expressed is GFP. GFP is 238 amino acids with a molecular weight of 27 kDa. A western blot will prove if GFP was successfully expressed after transfection. First, the transfected HeLa cells need to be lysed to break through the cell

membrane. Then, the HeLa cell debris is centrifuged to separate the proteins from the other cell components. The supernatant is tested with a western blot to locate a band at 27 kDa. If a band appears at 27 kDa, then this will confirm that GFP is expressed. Identifying GFP by fluorescence, flow cytometry, and western blot are ample evidence to conclude that *in vivo* protein ligation occurred. This evidence confirms that the experimental design works. This technique can be applied to pharmaceutical drugs. Sortase A can be used as a catalyst to trigger protein ligation between LPETG motif and Gly5 chain to activate expression of other proteins.

5.3 PRECISION MEDICINE AND SYNTHETIC CIRCUIT

This design pioneers synthetic biology by reengineering cellular pathways in a mammalian system. This technique uses Sortase A to carry out *in vivo* bioconjugation and bypass the ribosome. These synthetic cells will be accepted by the immune system [51, 52]. This can be applied in precision medicine as well. Precision medicine has four features that make it more advanced than regular medicine. Precision medicine is controlled in specificity, efficacy, selectivity, and time [53, 54]. In this design, precision medicine can be achieved by using synthetic circuits that control the Sortase A expression *in vivo*. Precision medicine is a specialized treatment to fit patients' needs. Every individual reacts differently to a particular disease. Therefore, precision medicine is able to treat a specific disease according to a patient's symptoms. It considers the intensity of a disease, environment, genetics, and lifestyle factors [55, 56]. This experiment has potential to apply Sortase A into precision medicine. Sortase A can control specificity by promoting protein ligation only between the LPETG motif and Gly5 chain. The efficiency is controlled by adding multiple binding domains. TOP1 and SH3 domains have a high binding affinity with each other, while BOT1 and SH3 domains have a low binding affinity with each other [38, 39]. So with the TOP1-SH3 and LPETG-GGGGG domains, the efficiency of protein ligation will increase. However, in the case of BOT1-SH3 and LPETG-GGGGG domains, the efficiency of protein ligation will decrease. Sortase A controls selectivity by expressing recombinant plasmids that contain the LPETG motif and Gly5 chain. Sortase A can distinguish between different drugs and activate specific medications. Finally, the expression of Sortase A can be controlled by synthetic circuits that makes it follow a biological clock. In general, these four features make Sortase A competent as precision medicine. This design uses synthetic circuits to catalyze *in vivo* ligation between a protein drug and the drug target.

REFERENCES

[1] Mazmanian, S. K., Liu, G., Ton-That, H., & Schneewind, O. (1999). Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science*, *285*(5428), 760–763. https://doi.org/10.1126/science.285.5428.760

[2] Huang, K. C., Mukhopadhyay, R., Wen, B., Gitai, Z., & Wingreen, N. S. (2008). Cell shape and cell-wall organization in Gram-negative bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19282–19287. https://doi.org/10.1073/pnas.0805309105

[3] Omeed Sizar, & Unakal, C. G. (2019, April). *Gram Positive Bacteria*. Nih.Gov; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK470553/

[4] Antos, J. M., Ingram, J., Fang, T., Pishesha, N., Truttmann, M. C., & Ploegh, H. L. (2017). Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation. *Current Protocols in Protein Science*, *89*, 15.3.1-15.3.19. https://doi.org/10.1002/cpps.38

[5] Pishesha, N., Ingram, J. R., & Ploegh, H. L. (2018). Sortase A: A Model for Transpeptidation and Its Biological Applications. *Annual Review of Cell and Developmental Biology*, *34*(1), 163–188. https://doi.org/10.1146/annurev-cellbio-100617-062527

[6] Mathur, M., Xiang, J. S., & Smolke, C. D. (2017). Mammalian synthetic biology for studying the cell. *Journal of Cell Biology*, *216*(1), 73–82. https://doi.org/10.1083/jcb.201611002

[7] Ding, Y., Wu, F., & Tan, C. (2014, December 19). Synthetic biology: A bridge between artificial and natural cells. *Life*. MDPI AG. https://doi.org/10.3390/life4041092

[8] Guimaraes, C. P., Witte, M. D., Theile, C. S., Bozkurt, G., Kundrat, L., Blom, A. E. M., & Ploegh, H. L. (2013). Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nature Protocols*, *8*(9), 1787–1799. https://doi.org/10.1038/nprot.2013.101

[9] Lu, C., Zhu, J., Wang, Y., Umeda, A., Cowmeadow, R. B., Lai, E., … Zhang, Z. (2007). Staphylococcus aureus sortase a exists as a dimeric protein in vitro. *Biochemistry*, *46*(32), 9346–9354. https://doi.org/10.1021/bi700519w

[10] Mazmanian, S. K., Ton-That, H., & Schneewind, O. (2001). Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus. *Molecular Microbiology*. https://doi.org/10.1046/j.1365-2958.2001.02411.x

[11] Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., & Ellestad, G. (2003). Kinetic mechanism of Staphylococcus aureus sortase SrtA. *Biochemistry*, *42*(38), 11307–11315. https://doi.org/10.1021/bi034391g

[12] Jacobitz, A. W., Kattke, M. D., Wereszczynski, J., & Clubb, R. T. (2017). Sortase Transpeptidases: Structural Biology and Catalytic Mechanism. In *Advances in Protein Chemistry and Structural Biology* (Vol. 109, pp. 223–264). Academic Press Inc. https://doi.org/10.1016/bs.apcsb.2017.04.008

[13] Zhu, J., Xiang, L., Jiang, F., & Zhang, Z. J. (2016). Equilibrium of sortase A dimerization on Staphylococcus aureus cell surface mediates its cell wall sorting activity. *Experimental Biology and Medicine*, *241*(1), 90–100. https://doi.org/10.1177/1535370215592122
[14] Nwe, K., & Brechbiel, M. W. (2009, June 1). Growing applications of "click chemistry" for bioconjugation in contemporary biomedical research. *Cancer Biotherapy and Radiopharmaceuticals*. https://doi.org/10.1089/cbr.2008.0626
[15] Meghani, N. M., Amin, H. H., & Lee, B. J. (2017, November 1). Mechanistic applications of click chemistry for pharmaceutical drug discovery and drug delivery. *Drug Discovery Today*. Elsevier Ltd. https://doi.org/10.1016/j.drudis.2017.07.007
[16] Hein, C. D., Liu, X. M., & Wang, D. (2008, October). Click chemistry, a powerful tool for pharmaceutical sciences. *Pharmaceutical Research*. https://doi.org/10.1007/s11095-008-9616-1
[17] Cañeque, T., Müller, S., & Rodriguez, R. (2018, September 1). Visualizing biologically active small molecules in cells using click chemistry. *Nature Reviews Chemistry*. Nature Publishing Group. https://doi.org/10.1038/s41570-018-0030-x
[18] Peplow, M. (2019). Click chemistry targets antibody-drug conjugates for the clinic. *Nature Biotechnology*. https://doi.org/10.1038/d41587-019-00017-4
[19] Meng, G., Guo, T., Ma, T., Zhang, J., Shen, Y., Sharpless, K. B., & Dong, J. (2019). Modular click chemistry libraries for functional screens using a diazotizing reagent. *Nature*, *574*(7776), 86–89. https://doi.org/10.1038/s41586-019-1589-1
[20] Pritz, S., Wolf, Y., Kraetke, O., Klose, J., Bienert, M., & Beyermann, M. (2007). Synthesis of Biologically Active Peptide Nucleic Acid−Peptide Conjugates by Sortase-Mediated Ligation. *The Journal of Organic Chemistry*, *72*(10), 3909–3912. https://doi.org/10.1021/jo062331l
[21] Patterson, D., Schwarz, B., Avera, J., Western, B., Hicks, M., Krugler, P., … Douglas, T. (2017). Sortase-Mediated Ligation as a Modular Approach for the Covalent Attachment of Proteins to the Exterior of the Bacteriophage P22 Virus-like Particle. *Bioconjugate Chemistry*, *28*(8), 2114–2124. https://doi.org/10.1021/acs.bioconjchem.7b00296
[22] Chen, L., Cohen, J., Song, X., Zhao, A., Ye, Z., Feulner, C. J., … Chen, P. R. (2016). Improved variants of SrtA for site-specific conjugation on antibodies and proteins with high efficiency. *Scientific Reports*, *6*. https://doi.org/10.1038/srep31899
[23] Bentley, M. L., Gaweska, H., Kielec, J. M., & McCafferty, D. G. (2007). Engineering the substrate specificity of Staphylococcus aureus sortase A: The β6/β7 loop from SrtB confers npqtn recognition to SrtA. *Journal of Biological Chemistry*, *282*(9), 6571–6581. https://doi.org/10.1074/jbc.M610519200
[24] Koussa, M. A., Sotomayor, M., & Wong, W. P. (2014). Protocol for sortase-mediated construction of DNA-protein hybrids and functional nanostructures. *Methods*, *67*(2), 134–141. https://doi.org/10.1016/j.ymeth.2014.02.020
[25] Jeong, H. J., Abhiraman, G. C., Story, C. M., Ingram, J. R., & Dougan, S. K. (2017). Generation of Ca2+-independent sortase A mutants with enhanced activity for protein and cell surface labeling. *PLoS ONE*, *12*(12). https://doi.org/10.1371/journal.pone.0189068

[26] Dorr, B. M., Ham, H. O., An, C., Chaikof, E. L., & Liu, D. R. (2014). Reprogramming the specificity of sortase enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(37), 13343–13348. https://doi.org/10.1073/pnas.1411179111
[27] Tsukiji, S., & Nagamune, T. (2009, March 23). Sortase-mediated ligation: A gift from gram-positive bacteria to protein engineering. *ChemBioChem*. https://doi.org/10.1002/cbic.200800724
[28] Li, J., Zhang, Y., Soubias, O., Khago, D., Chao, F. A., Li, Y., … Byrd, R. A. (2020). Optimization of sortase A ligation for flexible engineering of complex protein systems. *Journal of Biological Chemistry*, *295*(9), 2664–2675. https://doi.org/10.1074/jbc.RA119.012039
[29] Dai, X., Böker, A., & Glebe, U. (2019). Broadening the scope of sortagging. *RSC Advances*. Royal Society of Chemistry. https://doi.org/10.1039/c8ra06705h
[30] Lienert, F., Lohmueller, J. J., Garg, A., & Silver, P. A. (2014, February). Synthetic biology in mammalian cells: Next generation research tools and therapeutics. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm3738
[31] Bower, A. G., McClintock, M. K., & Fong, S. S. (2010). Synthetic biology: A foundation for multi-scale molecular biology. *Bioengineered Bugs*, *1*(5), 309–312. https://doi.org/10.4161/bbug.1.5.12391
[32] Serrano, L. (2007). Synthetic biology: Promises and challenges. *Molecular Systems Biology*. https://doi.org/10.1038/msb4100202
[33] Kojima, R., & Fussenegger, M. (2019, April 15). Synthetic Biology: Engineering Mammalian Cells To Control Cell-to-Cell Communication at Will. *ChemBioChem*. Wiley-VCH Verlag. https://doi.org/10.1002/cbic.201800682
[34] Buddingh', B. C., & Van Hest, J. C. M. (2017). Artificial Cells: Synthetic Compartments with Life-like Functionality and Adaptivity. *Accounts of Chemical Research*, *50*(4), 769–777. https://doi.org/10.1021/acs.accounts.6b00512
[35] Porcar, M., Danchin, A., de Lorenzo, V., dos Santos, V. A., Krasnogor, N., Rasmussen, S., & Moya, A. (2011). The ten grand challenges of synthetic life. *Systems and Synthetic Biology*, *5*(1), 1–9. https://doi.org/10.1007/s11693-011-9084-5
[36] Xu, C., Hu, S., & Chen, X. (2016, November 1). Artificial cells: from basic science to applications. *Materials Today*. Elsevier B.V. https://doi.org/10.1016/j.mattod.2016.02.020
[37] Kurochkina, N., & Guha, U. (2013). SH3 domains: Modules of protein-protein interactions. *Biophysical Reviews*. Springer Verlag. https://doi.org/10.1007/s12551-012-0081-z
[38] Wright, C. M., Van Der Merwe, M., DeBrot, A. H., & Bjornsti, M. A. (2015). DNA topoisomerase I domain interactions impact enzyme activity and sensitivity to camptothecin. *Journal of Biological Chemistry*, *290*(19), 12068–12078. https://doi.org/10.1074/jbc.M114.635078
[39] Mullaney, B. P., Pallavicini, M. G., & Marks, J. D. (2001). Epitope mapping of neutralizing botulinum neurotoxin A antibodies by phage display. *Infection and Immunity*, *69*(10), 6511–6514. https://doi.org/10.1128/IAI.69.10.6511-6514.2001

[40] Umeda, A., Thibodeaux, G. N., Moncivais, K., Jiang, F., & Zhang, Z. J. (2010). A versatile approach to transform low-affinity peptides into protein probes with cotranslationally expressed chemical cross-linker. *Analytical Biochemistry*, *405*(1), 82–88. https://doi.org/10.1016/j.ab.2010.05.026
[41] Chiang, C. F., Okou, D. T., Griffin, T. B., Verret, C. R., & Williams, M. N. V. (2001). Green fluorescent protein rendered susceptible to proteolysis: Positions for protease-sensitive insertions. *Archives of Biochemistry and Biophysics*, *394*(2), 229–235. https://doi.org/10.1006/abbi.2001.2537
[42] Svalastog, A. L., & Martinelli, ucia. (2013). Representing life as opposed to being: The bio-objectification process of the hela cells and its relation to personalized medicine. *Croatian Medical Journal*, *54*(4), 397–402. https://doi.org/10.3325/cmj.2013.54.397
[43] Landry, J. J. M., Pyl, P. T., Rausch, T., Zichner, T., Tekkedil, M. M., Stütz, A. M., … Steinmetz, L. M. (2013). The genomic and transcriptomic landscape of a hela cell line. *G3: Genes, Genomes, Genetics*, *3*(8), 1213–1224. https://doi.org/10.1534/g3.113.005777
[44] Yang, Y., Song, H., He, D., Zhang, S., Dai, S., Xie, X., … Chen, P. R. (2017). Genetically encoded releasable photo-cross-linking strategies for studying protein-protein interactions in living cells. *Nature Protocols*, *12*(10), 2147–2168. https://doi.org/10.1038/nprot.2017.090
[45] Wingfield, P. T. (2017). N-terminal methionine processing. *Current Protocols in Protein Science*, *2017*. https://doi.org/10.1002/cpps.29
[46] Liao, Y.-D., Jeng, J.-C., Wang, C.-F., Wang, S.-C., & Chang, S.-T. (2004). Removal of N-terminal methionine from recombinant proteins by engineered E. coli methionine aminopeptidase . *Protein Science*, *13*(7), 1802–1810. https://doi.org/10.1110/ps.04679104
[47] Xiao, Q., Zhang, F., Nacev, B. A., Liu, J. O., & Pei, D. (2010). Protein N-terminal processing: Substrate specificity of escherichia coli and human methionine aminopeptidases. *Biochemistry*, *49*(26), 5588–5599. https://doi.org/10.1021/bi1005464
[48] Km, M. (2018, February 21). *Flow Cytometry: An Overview*. Current Protocols in Immunology. https://pubmed.ncbi.nlm.nih.gov/29512141/
[49] Adan, A., Alizada, G., Kiraz, Y., Baran, Y., & Nalbant, A. (2017, February 17). Flow cytometry: basic principles and applications. *Critical Reviews in Biotechnology*. Taylor and Francis Ltd. https://doi.org/10.3109/07388551.2015.1128876
[50] Rubio, F. J., Li, X., Liu, Q. R., Cimbro, R., & Hope, B. T. (2016). Fluorescence activated cell sorting (FACS) and gene expression analysis of fos-expressing neurons from fresh and frozen rat brain tissue. *Journal of Visualized Experiments*, *2016*(114). https://doi.org/10.3791/54358
[51] Ginsburg, G. S., & Phillips, K. A. (2018). Precision medicine: From science to value. *Health Affairs*, *37*(5), 694–701. https://doi.org/10.1377/hlthaff.2017.1624
[52] Ir, K., O, F., G, H., E, von M., & Mv, K. (2017, October 19). *What Is Precision Medicine?* The European Respiratory Journal. https://pubmed.ncbi.nlm.nih.gov/29051268/

[53] Fox, B. I., & Felkey, B. G. (2016). Precision medicine: Considerable potential combined with important challenges. *Hospital Pharmacy*, *51*(11), 950–951. https://doi.org/10.1310/hpj5111-950
[54] Way, J. C., Collins, J. J., Keasling, J. D., & Silver, P. A. (2014, March 27). Integrating biological redesign: Where synthetic biology came from and where it needs to go. Cell. Cell Press. https://doi.org/10.1016/j.cell.2014.02.039
[55] Pinho, J. R. R. (2017, January 1). Precision Medicine. *Einstein (Sao Paulo, Brazil)*. https://doi.org/10.1590/S1679-45082017ED4016
[56] Krzyszczyk, P., Acevedo, A., Davidoff, E. J., Timmins, L. M., Marrero-Berrios, I., Patel, M., … Yarmush, M. L. (2018). The growing role of precision and personalized medicine for cancer treatment. *TECHNOLOGY*, *06*(03n04), 79–100. https://doi.org/10.1142/s2339547818300020

APPENDIX A: GFP GENE SEQUENCE

The DNA sequence is 717 base pairs and 238 amino acids. This GFP gene is from SnapGene (version 5.1).

ATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGAT
GGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAAC
ATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATG
GCCAACACTTGTCACTACTTTCTCTTATGGTGTTCAATGCTTTTCAAGATACCCAGAT
CATATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAA
AGAACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTT
GAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGAT
GGAAACATTCTTGGACACAAATTGGAATACAACTATAACTCACACAATGTATACATC
ATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAAAATTAGACACAACAT
TGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCG
ATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAA
AGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGG
GATTACACATGGCATGGATGAACTATACAAATAG

APPENDIX B: RECOMBINANT PLASMID 1 SEQUENCE

Recombinant plasmid 1 was designed on ApE (version 2.0.61) with the pEF-GFP vector backbone from Addgene. The total length of plasmid 1 is 6353 nucleotides.

GTCGACATTGATTATTGACTAGATCATCGCGTGAGGCTCCGGTGCCCGTCAGTGGGC
AGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAA
CCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGG
CTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTG
AACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGT
TCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCA
CGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGG
GAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCC
TGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTC
GCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTT
TTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGG
TTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCG
AGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGG
CCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGC
AAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCC
CTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGA
GTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTC
CACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTAC

GTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTG
GGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGC
CCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTT
TTTCTTCCATTTCAGGTGTCGTGAGGAATTCGCCACCATGAGTAAAGGAGAAGAACT
TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAA
ATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAA
ATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTC
TCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACGGCATGACTTTT
TCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATG
ACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAAT
AGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAA
ATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAGACT
ACAAAGACGATGATGACAAACTGCCTGAAACCGGTTAATTAATAAAGGAAATTTAT
TTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCAGGTACCGCTCCGGTGCCC
GTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTC
GGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGT
CGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGT
AGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGC
CGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGA
ATTACTTCCACGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGG
AAGTGGGTGGGAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTG
AGTTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTC
GCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGC
TGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACAC
TGGTATTTCGGTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCAC
ATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGT
CTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGC
CCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCG
CTTCCCGGCCCTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCG
GGCGGGTGAGTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTT
CATGTGACTCCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCT
TTTGGAGTACGTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCC
ACACTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCT
TGGAATTTGCCCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGT
TCAAAGTTTTTTTCTTCCATTTCAGGTGTCGTGAGCTAGCCACCATGGGAGGTGGAG
GTGGAGACTACAAAGACGATGATGACAAGAAGAATGGAATCAAAGTTAACTTCAAA
ATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAA
TACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCCACACA
ATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAGTT

TGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTATACAAATAAAGCGGCC
GCACTCCTCAGGTGCAGGCTGCCTATCAGAAGGTGGTGGCTGGTGTGGCCAATGCCC
TGGCTCACAAATACCACTGAGATCTTTTTCCCTCTGCCAAAAATTATGGGGACATCA
TGAAGCCCCTTGAGCATCTGACTTCTGGCTAATAAAGGAAATTTATTTTCATTGCAA
TAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGGAAGGACATATGGGAGGGCAAATC
ATTTAAAACATCAGAATGAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGC
TGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCC
CCTGCTGTCCATTCCTTATTCCATAGAAAAGCCTTGACTTGAGGTTAGATTTTTTTTA
TATTTTGTTTTGTGTTATTTTTTTCTTTAACATCCCTAAAATTTTCCTTACATGTTTTAC
TAGCCAGATTTTTCCTCCTCTCCTGACTACTCCCAGTCATAGCTGTCCCTCTTCTCTTA
TGGAGATCCCTCGACCTGCAGCCCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAA
GTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTC
ACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCGGATCCGCATCTCAATTAG
TCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGT
TCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGG
CCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAG
GCTTTTGCAAAAAGCTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAA
TAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTG
TCCAAACTCATCAATGTATCTTATCATGTCTGGATCCGCTGCATTAATGAATCGGCC
AACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTG
ACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAA
GGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAG
GCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAA
ACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCT
CTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAG
CGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGC
TCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCC
GGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT
GAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCT
GCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAA
CCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAA
AAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACG
AAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGA
TCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTAT
TTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGG

GCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTC
CAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCT
GCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGT
AGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTG
TCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGA
GTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATC
GTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCAT
AATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAA
CCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAA
TACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATG
TAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
GGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACG
GAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGT
TATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGG
GTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGG

APPENDIX C: RECOMBINANT PLASMID 2 SEQUENCE

Recombinant plasmid 2 was designed on ApE (version 2.0.61) with the pEF-GFP vector backbone from Addgene. The total length of plasmid 2 is 6548 nucleotides. This plasmid includes TOP1 and SH3 domains that will increase the expression of GFP.

GTCGACATTGATTATTGACTAGATCATCGCGTGAGGCTCCGGTGCCCGTCAGTGGGC
AGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAA
CCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGG
CTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTG
AACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGT
TCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCA
CGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGG
GAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCC
TGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTC
GCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTT
TTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGG
TTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCG
AGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGG
CCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGC
AAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCC
CTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGA
GTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTC

CACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTAC
GTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTG
GGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGC
CCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTT
TTTCTTCCATTTCAGGTGTCGTGAGGAATTCGCCACCATGAGTAAAGGAGAAGAACT
TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAA
ATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAA
ATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTC
TCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACGGCATGACTTTT
TCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATG
ACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAAT
AGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAA
ATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAGACT
ACAAAGACGATGATGACAAAGCTCCAGCCTTTCCACCTCCTAGCCCTCCACTGCCTG
AAACCGGTTAATTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTT
TGTGTCTCTCAGGTACCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCAC
AGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGT
GGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGG
GTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACG
GGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCT
TTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCACGCCCCTGGCTGCAGTACG
TGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGGGAGAGTTCGAGGCCTTGCG
CTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCCTGGCCTGGGCGCTGGGGCC
GCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTCGCTGCTTTCGATAAGTCTCT
AGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTT
GTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGGTTTTTGGGGCCGCGGGCGGC
GACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCG
GCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTG
GCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGGCA
CCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCAAA
ATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAA
AGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCG
TCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGGGG
GAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAGTTAG
GCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTTGGATCT
TGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTCAGGTGT
CGTGAGCTAGCCACCATGGGAGGTGGAGGTGGAGACTACAAAGACGATGATGACAA
GAATGATCCTAACCTCTTTGTCGCACTGTATGATTTCGTCGCCTCTGGTGACAATACC
CTCAGCATTACTAAAGGAGAGAAGCTGAGGGTGCTGGGATACAATCACAACGGTGA

GTGGTGCGAGCCTCAGACAAAGAACGGACAGGGCTGGGTCCCTTCCAACTACAAGA
ATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTCAA
CTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCA
GACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAG
AGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGA
TGAACTATACAAATAAAGCGGCCGCACTCCTCAGGTGCAGGCTGCCTATCAGAAGG
TGGTGGCTGGTGTGGCCAATGCCCTGGCTCACAAATACCACTGAGATCTTTTTCCCT
CTGCCAAAAATTATGGGGACATCATGAAGCCCCTTGAGCATCTGACTTCTGGCTAAT
AAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGG
AAGGACATATGGGAGGGCAAATCATTTAAAACATCAGAATGAGTATTTGGTTTAGA
GTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAGAG
GTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGCCT
TGACTTGAGGTTAGATTTTTTTTATATTTTGTTTTGTGTTATTTTTTTCTTTAACATCCC
TAAAATTTTCCTTACATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTGACTACTCCCA
GTCATAGCTGTCCCTCTTCTCTTATGGAGATCCCTCGACCTGCAGCCCAAGCTTGGC
GTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC
AACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTA
ACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG
CCAGCGGATCCGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCC
ATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTT
TTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGT
GAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTAACTTGTTTATTGCAGC
TTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTT
TTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGG
ATCCGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGG
CGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAG
CGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC
GCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGG
CCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC
GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACC
TGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT
TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGT
ATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA
GAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG
GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCA
AGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA

CGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGA
TTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCA
ATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAG
GCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCG
TGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATAC
CGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGA
AGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAAT
TGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTT
GCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCT
CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCG
GTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA
CTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCT
TTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
CGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC
CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCAT
CTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA
AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAAT
ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT
TTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTG
G

APPENDIX D: RECOMBINANT PLASMID 3 SEQUENCE

Recombinant plasmid 3 was designed on ApE (version 2.0.61) with the pEF-GFP vector backbone from Addgene. The total length of plasmid 3 is 6548 nucleotides. This plasmid includes BOT1 and SH3 domains that will decrease the expression of GFP.

GTCGACATTGATTATTGACTAGATCATCGCGTGAGGCTCCGGTGCCCGTCAGTGGGC
AGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAA
CCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGG
CTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTG
AACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGT
TCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCA
CGCCCCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGG
GAGAGTTCGAGGCCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCC
TGGCCTGGGCGCTGGGGCCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTC
GCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTT
TTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGG
TTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCG

AGGCGGGGCCTGCGAGCGCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGG
CCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGC
AAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCC
CTGCTGCAGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGA
GTCACCCACACAAAGGAAAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTC
CACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTAC
GTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTG
GGTGGAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGC
CCTTTTTGAGTTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTT
TTTCTTCCATTTCAGGTGTCGTGAGGAATTCGCCACCATGAGTAAAGGAGAAGAACT
TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAA
ATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTACCCTTAA
ATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTTTC
TCTTATGGTGTTCAATGCTTTTCAAGATACCCAGATCATATGAAACGGCATGACTTTT
TCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATG
ACGGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAAT
AGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTTGGACACAA
ATTGGAATACAACTATAACTCACACAATGTATACATCATGGCAGACAAACAAGACT
ACAAAGACGATGATGACAAAGCTGCTGCAATGCAGAAACCTAGCCTGCCACTGCCT
GAAACCGGTTAATTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTT
TTTGTGTCTCTCAGGTACCGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCC
ACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAG
GTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGA
GGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAA
CGGGTTTGCCGCCAGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCT
CTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTTCCACGCCCCTGGCTGCAGTA
CGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGGGAGAGTTCGAGGCCTTG
CGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCCTGGCCTGGGCGCTGGGG
CCGCCGCGTGCGAATCTGGTGGCACCTTCGCGCCTGTCTCGCTGCTTTCGATAAGTC
TCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTTTTTTCTGGCAAGATAGT
CTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGGTTTTTGGGGCCGCGGGC
GGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGC
GCGGCCACCGAGAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGC
CTGGCCTCGCGCCGCCGTGTATCGCCCCGCCCTGGGCGGCAAGGCTGGCCCGGTCGG
CACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTGCAGGGAGCTCA
AAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGA
AAAGGGCCTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGC
CGTCCAGGCACCTCGATTAGTTCTCGAGCTTTTGGAGTACGTCGTCTTTAGGTTGGG
GGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAGTT

AGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTTGGAT
CTTGGTTCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTCAGGT
GTCGTGAGCTAGCCACCATGGGAGGTGGAGGTGGAGACTACAAAGACGATGATGAC
AAGAATGATCCTAACCTCTTTGTCGCACTGTATGATTTCGTCGCCTCTGGTGACAATA
CCCTCAGCATTACTAAAGGAGAGAAGCTGAGGGTGCTGGGATACAATCACAACGGT
GAGTGGTGCGAGCCTCAGACAAAGAACGGACAGGGCTGGGTCCCTTCCAACTACAA
GAATGGAATCAAAGTTAACTTCAAAATTAGACACAACATTGAAGATGGAAGCGTTC
AACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTAC
CAGACAACCATTACCTGTCCACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAG
AGAGACCACATGGTCCTTCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATG
GATGAACTATACAAATAAAGCGGCCGCACTCCTCAGGTGCAGGCTGCCTATCAGAA
GGTGGTGGCTGGTGTGGCCAATGCCCTGGCTCACAAATACCACTGAGATCTTTTTCC
CTCTGCCAAAAATTATGGGGACATCATGAAGCCCCTTGAGCATCTGACTTCTGGCTA
ATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTC
GGAAGGACATATGGGAGGGCAAATCATTTAAAACATCAGAATGAGTATTTGGTTTA
GAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAG
AGGTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGC
CTTGACTTGAGGTTAGATTTTTTTTATATTTTGTTTTGTGTTATTTTTTTCTTTAACATC
CCTAAAATTTTCCTTACATGTTTTACTAGCCAGATTTTTCCTCCTCTCCTGACTACTCC
CAGTCATAGCTGTCCCTCTTCTCTTATGGAGATCCCTCGACCTGCAGCCCAAGCTTG
GCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCAC
ACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGC
TAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT
GCCAGCGGATCCGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCC
CATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATT
TTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAG
TGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTAACTTGTTTATTGCAG
CTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTT
TTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTG
GATCCGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGG
GCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGA
GCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAA
CGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAG
GCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAAT
CGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTT
TCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATAC
CTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGT
ATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCG
TTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAA

GACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGA
AGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTT
GGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGC
AAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCT
ACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAG
ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATC
AATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGA
GGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTC
GTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATA
CCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGG
AAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAA
TTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGT
TGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGC
TCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCG
GTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA
CTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCT
TTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC
CGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT
TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC
CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCAT
CTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA
AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAAT
ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT
TTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTG
G

APPENDIX E: RECOMBINANT PLASMID 4 SEQUENCE

Recombinant plasmid 4 encodes for Sortase A. Sortase A gene is 621 base pairs and the protein sequence is 206 amino acids. The total length of this plasmid is 710 base pairs.

NTGTCGTGAGGATTAGCTTGGTACTAATACGACTCACTATAGGGAGACCCAAGCTGG
CTAGGTAAGCTTGGTACCGAGCTCGGATCCACTAGTATGGGCCAAGCTAAACCTCAA
ATTCCGAAAGATAAATCAAAAGTGGCAGGCTATATTGAAATTCCAGATGCTGATATT
AAAGAACCAGTATATCCAGGACCAGCAACACCTGAACAATTAAATAGAGGTGTAAG
CTTTGCAGAAGAAAATGAATCACTAGATGATCAAAATATTTCAATTGCAGGACACA
CTTTCATTGACCGTCCGAACTATCAATTTACAAATCTTAAAGCAGCCAAAAAAGGTA
GTATGGTGTACTTTAAAGTTGGTAATGAAACACGTAAGTATAAAATGACAAGTATA
AGAGGTGTTAAGCCAACAGATGTAGAAGTTCTAGATGAACAAAAAGGTAAAGATAA

ACAATTAACATTAATTACTTGTGATGATTACAATGAAAAGACAGGCGTTTGGGAAA
AACGTAAAATCTTTGTAGCTACAGAAGTCAAAGAATTCTGCAGATATCCAGCACAGT
GGCGGCCGCTCGAGTCTAGAGGGCCCTTCGAACAAAAACTCATCTCAGAAGAGGAT
CTGAATATGCATACCGGTCATCATCACCATCACCATTGAGTTTAAACCCGCTGATCA
GCCTCGACTGTGCCTTCTAGTTGCCAGCCATC